PROBABILITY AND ITS ENGINEERING USES. *By* THORNTON C. FRY.

ELEMENTARY DIFFERENTIAL EQUATIONS. *By* THORNTON C. FRY.

TRANSMISSION CIRCUITS FOR TELEPHONIC COMMUNICATION. METHODS OF ANALYSIS AND DESIGN. *By* K. S. JOHNSON.

TRANSMISSION NETWORKS AND WAVE FILTERS. *By* T. E. SHEA.

ECONOMIC CONTROL OF QUALITY OF MANUFACTURED PRODUCT. *By* W. A. SHEWHART.

ELECTROMECHANICAL TRANSDUCERS AND WAVE FILTERS. *By* WARREN P. MASON. Second Edition.

RHOMBIC ANTENNA DESIGN. *By* A. E. HARPER.

POISSON'S EXPONENTIAL BINOMIAL LIMIT. *By* E. C. MOLINA.

ELECTROMAGNETIC WAVES. *By* S. A. SCHELKUNOFF.

NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN. *By* HENDRICK W. BODE.

QUARTZ CRYSTALS FOR ELECTRICAL CIRCUITS. *By* R. A. HEISING.

CAPACITORS—THEIR USE IN ELECTRONIC CIRCUITS. *By* M. BROTHERTON.

FOURIER INTEGRALS FOR PRACTICAL APPLICATIONS. *By* GEORGE A. CAMPBELL AND RONALD M. FOSTER.

VISIBLE SPEECH. *By* RALPH K. POTTER, GEORGE A. KOPP, AND HARRIET C. GREEN.

APPLIED MATHEMATICS FOR ENGINEERS AND SCIENTISTS. *By* S. A. SCHELKUNOFF.

EARTH CONDUCTION EFFECTS IN TRANSMISSION SYSTEMS. *By* ERLING D. SUNDE.

THEORY AND DESIGN OF ELECTRON BEAMS. *By* J. R. PIERCE. Second Edition.

PIEZOELECTRIC CRYSTALS AND THEIR APPLICATION TO ULTRASONICS. *By* WARREN P. MASON.

MICROWAVE ELECTRONICS. *By* JOHN C. SLATER.

PRINCIPLES AND APPLICATIONS OF WAVEGUIDE TRANSMISSION. *By* GEORGE C. SOUTHWORTH.

TRAVELING-WAVE TUBES. *By* J. R. PIERCE.

ELECTRONS AND HOLES IN SEMICONDUCTORS. *By* WILLIAM SHOCKLEY.

FERROMAGNETISM. *By* RICHARD M. BOZORTH.

THE DESIGN OF SWITCHING CIRCUITS. *By* WILLIAM KEISTER, ALASTAIR E. RITCHIE, AND SETH H. WASHBURN.

SPEECH AND HEARING IN COMMUNICATION. *By* HARVEY FLETCHER. Second Edition.

MODULATION THEORY. *By* HAROLD S. BLACK.

SWITCHING RELAY DESIGN. *By* R. L. PEEK, JR. AND H. N. WAGAR.

APPLIED MATHEMATICS
FOR
ENGINEERS AND SCIENTISTS

S. A. SCHELKUNOFF
MEMBER OF THE TECHNICAL STAFF OF
BELL TELEPHONE LABORATORIES, INC.

FOURTH PRINTING

D. VAN NOSTRAND COMPANY, INC.
PRINCETON, NEW JERSEY
TORONTO LONDON
NEW YORK

D. VAN NOSTRAND COMPANY, INC.
120 Alexander St., Princeton, New Jersey (*Principal office*)
257 Fourth Avenue, New York 10, New York

D. VAN NOSTRAND COMPANY, LTD.
358, Kensington High Street, London, W.14, England

D. VAN NOSTRAND COMPANY (Canada), LTD.
25 Hollinger Road, Toronto 16, Canada

First Published November 1948
Reprinted May 1951, May 1955,
November 1957

Produced by
TECHNICAL COMPOSITION CO.
BOSTON, MASS.

PRINTED IN THE UNITED STATES OF A

PREFACE

Mathematics is playing an increasingly important role in the development of science and engineering. Books on "Higher Mathematics for Engineers" which met the needs of engineers only a short time ago are no longer adequate. Without more advanced mathematical knowledge, many recent achievements cannot be fully understood. Instructors in graduate scientific and engineering courses find themselves increasingly forced either to take time from the main subject in order to develop needed mathematics, or to restrict the scope of instruction. The purpose of this book is to bring "Higher Mathematics for Scientists and Engineers" up to date. The topics selected are those most frequently useful in various fields. Topics that are of interest primarily in specialized fields or that are only occasionally useful are treated lightly, if at all. Simple physical illustrations have been used, whenever helpful, to explain mathematical ideas and methods; but examples requiring more specialized knowledge have been avoided, for they might be confusing to some, even though helpful to others. In general, the book is devoted to those branches of mathematics which are needed in mathematical physics and engineering.

The book is divided essentially into two parts, one devoted to general mathematical methods and the other to special transcendental functions. The topics in the first part are arranged in the order of their difficulty. Thus the more elementary applications of Functions of a Complex Variable are considered in the first chapter. This chapter serves also as an introduction to the later and more advanced chapters on this subject. The second chapter is an introduction to the Theory of Approximation; it contains Newton's and Lagrange's interpolation formulas, it explains the basic difference between the approximations in the vicinity of a point by power series and approximations on the average by series of Legendre polynomials, and it introduces Fourier series. Chapter 3 is devoted to two general methods of solving algebraic and transcendental equations: the graphical or table method and the straightforward method of perturbations or successive approximations. Power series are considered in Chapter 4, general ideas of differentiation and integration in Chapters 5 and 6. While a great deal of Vector Analysis is needed in some fields, such as mechanics, only the more elementary ideas are most frequently useful; and these are treated in Chapter 7. Chapter 8 on the important coordinate systems is primarily for reference. In Chapter 9 exponential and loga-

rithmic functions are used as examples to explain the method of defining functions as solutions of differential equations. This concludes the more elementary part of the book.

The next three chapters are devoted to ordinary linear differential equations: Chapter 10 to those of the first order, Chapter 11 to those of the second order, and Chapter 12 to those of higher orders. Because of their importance in the theories of oscillations and waves, the second order equations are treated with particular thoroughness. Several sections of this chapter are devoted to the "Wave Perturbation Method" which is the most important practical method of obtaining analytic approximations to the solutions of these equations. The basic idea behind this method is that the solutions of second order differential equations with variable coefficients are distorted or "perturbed" sinusoidal or exponential functions. The well-known WKB approximation is the first in a series of approximations which can be obtained by this method. Other sections in this chapter are devoted to orthogonal properties of the solutions of differential equations and to orthogonal expansions. Conformal mapping, with applications to the solution of electrostatic problems and problems of flow, and contour integration are discussed in Chapters 14 and 15. Chapter 16 is an introduction to the methods of treatment of linear dynamical systems, particularly to the Laplace Transform Method.

All the remaining chapters but the last are devoted to special functions: Gamma Functions, Sine and Cosine Integrals, Fresnel Integrals, Bessel Functions, and Legendre Functions. The Sine and Cosine Integrals have assumed a new importance in the solution of second order differential equations, in addition to the indispensable part played by them in the theory of radiation. In view of their increasing applications Bessel and Legendre Functions are treated much more thoroughly than has been customary in the past in books of this type. Functions of the second kind and functions of fractional degree and order are given equal status with the simpler functions of the first kind. The last chapter on Formulation of Equations is intended to connect this treatise on appliable mathematics with those on mathematical physics and mathematical engineering.

I am greatly indebted to Miss Marion C. Gray for constructive criticism and valuable suggestions during the preparation of this manuscript, and for her untiring efforts to insure its accuracy. I also wish to thank Miss Helene Sumoska who assisted in reading the proofs and in preparing the index. To Mr. B. A. Clarke the credit should go for advice in connection with the illustrations and to Mr. H. P. Gridley for the drawings.

S. A. S.

NEW YORK, N. Y.
May 17, 1948

CONTENTS

Chapter I. COMPLEX NUMBERS

Chapter II. THEORY OF APPROXIMATION

Chapter III. SOLUTION OF EQUATIONS

Chapter IV. POWER SERIES

Chapter V. DIFFERENTIATION

CHAPTER I

Complex Numbers

The range of applications of complex numbers is very wide. Some applications suggest themselves almost as soon as the meaning and the most elementary properties of these numbers are understood; others are more recondite, requiring an extensive knowledge of the theory of functions of a complex variable. Intimate familiarity with a comparatively few basic ideas is indispensable for success in practical applications of complex numbers; knowledge of formal rules alone is not particularly helpful. The only way to acquire this familiarity is to use the ideas again and again under a variety of circumstances. For this reason we shall consider a few geometric applications which are not of direct interest to physicists and engineers but which in the end will pay dividends.

Mathematicians, most physicists, and many engineers use "i" for the imaginary unit. Electrical engineers use "j"; but many confine this usage to elementary applications and then pass on to the older and more universal symbol "i." There is no alternative for electrical engineers but to accustom themselves to both notations so that they are neither annoyed nor impeded in their thinking when they encounter these conflicting symbols in various books and periodicals.

1. *Historical origin of complex numbers*

Complex numbers first arise in the course of solution of quadratic equations. Since the square of a real number, whether positive or negative, is positive, the equation

$$x^2 + 1 = 0 \tag{1}$$

does not appear to have any solution. Perhaps it has not; but if we postulate that this particular equation has two unknown roots, $\pm x$, we shall find that all other quadratic equations can be solved in terms of this "x." For instance, the roots of $y^2 + 4 = 0$ are $y = \pm 2x$. Similarly, the solution of

$$y^2 + 2y + 5 = 0 \tag{2}$$

is

$$(y + 1)^2 + 4 = 0, \qquad y + 1 = \pm 2x, \qquad y = -1 \pm 2x.$$

This is just a beginning. All algebraic equations can be solved in terms of x satisfying (1); they can be "solved" in the sense that if we substitute

for the unknown certain expressions of the form $a + bx$, where a and b are ordinary positive and negative numbers, and then perform the indicated algebraic operations in accordance with the rules of ordinary algebra, replacing $x^2 + 1$ by zero (or x^2 by -1), the result becomes an identity. For instance, substituting $-1 + 2x$ on the left side of (2), we have

$$\begin{aligned}(-1 + 2x)^2 + 2(-1 + 2x) + 5 &= 1 - 4x + 4x^2 - 2 + 4x + 5 \\ &= 1 - 4x - 4 - 2 + 4x + 5 = 0;\end{aligned}$$

hence equation (2) is " satisfied " by $-1 + 2x$.

The letter x is normally used for any unknown and some other letter is clearly desirable for our special purpose. In mathematics it is a universal custom to denote by " i " one of the roots of (1); the other is "$-i$" since $(-i)^2 = i^2$. The expression $a + ib$, where a and b are ordinary "real" numbers, is called a *complex number*; a is called the *real part* and b the *imaginary part* of the complex number. Symbolically we shall write

$$\operatorname{re}(a + ib) = a, \qquad \operatorname{im}(a + ib) = b. \tag{3}$$

2. *Algebra of complex numbers*

In performing various arithmetic operations on complex numbers the general rule is to treat complex numbers as ordinary algebraic expressions subject to the simplification $i^2 = -1$. The following is the summary of specific applications of this rule:

Addition

$$(x + iy) + (u + iv) = (x + u) + i(y + v). \tag{4}$$

Subtraction

$$(x + iy) - (u + iv) = (x - u) + i(y - v). \tag{5}$$

Multiplication

$$(x + iy)(u + iv) = (xu - yv) + i(yu + xv). \tag{6}$$

A complex number is equal to zero only if its real and imaginary parts are equal to zero separately; thus

$$\text{if} \quad x + iy = 0, \qquad \text{then} \quad x = y = 0. \tag{7}$$

To prove this we note that $(x + iy)(x - iy) = x^2 + y^2$. By hypothesis the first factor vanishes; hence the product also vanishes. Since x and y are real, x^2 and y^2 are positive; thus x and y must vanish.

Two complex numbers are equal if and only if their real and imaginary parts are equal; thus

$$\text{if} \quad x + iy = u + iv, \qquad \text{then} \quad x = u, \quad y = v. \tag{8}$$

This follows from (7).

Division

$$\frac{x + iy}{u + iv} = \frac{xu + yv}{u^2 + v^2} + i\frac{yu - xv}{u^2 + v^2}. \tag{9}$$

In order to obtain this result let the quotient be $a + ib$; by definition of the quotient*

$$(u + iv)(a + ib) = x + iy;$$

hence

$$(ua - vb) + i(va + ub) = x + iy,$$

$$ua - vb = x, \qquad va + ub = y.$$

It only remains to solve these equations for a and b.

Considering that "i" is defined arbitrarily as a root of equation (1), an immediate question may be asked: can we say that this equation possesses another root, "j" let us say, which is different from "i" and from "$-i$"? Having set the precedent by imagining a new number "i" to be a root of an equation which has no root by previous arithmetic standards, we are in no position to brush aside a positive answer. Gauss has shown, however, that if another complex unit is added, the number system does not obey all the laws of common algebra. We can answer our question in the affirmative provided we are willing to sacrifice some simple algebraic rule. Quaternions are numbers of the form $a + ib + jc + kd$, where $i^2 = j^2 = k^2 = -1$; these numbers violate the commutative law of multiplication. The development of an algebra of quaternions is a concrete reminder of a very important point: it is not enough to assume that complex numbers obey the laws of ordinary algebra; we have to prove it.

Suppose, for instance, we decide to introduce a quantity "j" $\neq \pm 1$ but subject to the condition $j^2 = 1$. Suppose we wish to perform algebraic operations on these new "complex numbers," $x + jy$, in accordance with the laws of common algebra. We multiply

$$(x + jy)(x - jy) = x^2 - jxy + jyx - j^2y^2 = x^2 - y^2$$

and find that the product can vanish when $y = \pm x$ without either of the factors being equal to zero. This result is contrary to what we are accustomed to expect in common algebra.

* This method is based directly on the definition of the quotient of two numbers. A simpler method of obtaining (9) is to multiply the numerator and denominator on the left side by $u - iv$.

Problems

1. Show that

a) $(x + iy)^2 = (x^2 - y^2) + 2ixy,$

b) $(x - iy)^3 = (x^3 - 3xy^2) - i(3x^2y - y^3),$

c) $(x + iy)(x - iy) = x^2 + y^2,$

d) $(x + iy)^2(x - iy)^2 = (x^2 + y^2)^2,$

e) $(x + iy)^n(x - iy)^n = (x^2 + y^2)^n.$

2. Show that

a) $\dfrac{1}{x - iy} = \dfrac{x + iy}{x^2 + y^2},$ b) $\dfrac{1}{(x + iy)^2} = \dfrac{(x^2 - y^2) - 2ixy}{(x^2 + y^2)^2}.$

3. Show that

$$\sqrt{x + iy} = \pm\left[\sqrt{\tfrac{1}{2}(\sqrt{x^2 + y^2} + x)} + i\sqrt{\tfrac{1}{2}(\sqrt{x^2 + y^2} - x)}\right], \quad \text{if } y > 0;$$

$$= \pm\left[\sqrt{\tfrac{1}{2}(\sqrt{x^2 + y^2} + x)} - i\sqrt{\tfrac{1}{2}(\sqrt{x^2 + y^2} - x)}\right], \quad \text{if } y < 0.$$

Hint: Let $\sqrt{x + iy} = u + iv$, square and solve for u and v. Give reasons f rejecting some solutions and for the difference in signs preceding "i" in the two case

3. *Laws of common algebra*

The following are the rules or " laws " or " postulates " of manipulatio which are characteristic of common algebra.

Commutative law

for addition: $a + b = b + a;$

for multiplication: $ab = ba.$

Associative law

for addition: $a + (b + c) = (a + b) + c;$

for multiplication: $a(bc) = (ab)c.$

Distributive law

$$a(b + c) = ab + ac.$$

Identity elements

for addition the identity element is called " zero "

$$a + 0 = a,$$

for multiplication the identity element is called " unity "

$$a \cdot 1 = a.$$

Additive inverse

there is an element $(-a)$ such that

$$a + (-a) = 0.$$

Cancellation law

$$\text{If} \quad c \neq 0 \quad \text{and} \quad ac = bc, \qquad \text{then} \quad a = b.$$

The same result is obtained if this law is formulated as follows:

$$\text{if} \quad ab = 0, \qquad \text{then either} \quad a = 0 \quad \text{or} \quad b = 0;$$

that is, the product does not vanish unless one of its factors vanishes.

There exist proofs that the above laws are obeyed by integers, fractions, all real numbers, *and* complex numbers as defined in the preceding sections. No further extension of the number system is possible without sacrificing at least one of these rules of manipulation. In the algebra of vectors the cancellation law does not hold, the commutative law holds for scalar multiplication but not for vector multiplication — to cite but two examples.

4. *Complex numbers as kinematic operators*

From a logical point of view mathematics is concerned not with substance but with form, not with things but with relations between them. It is this attribute of mathematics that makes its applications possible. Relations within one set of things may in some respects be the same as relations within another set. Expressions $a + bi$, where i is a particular root of (1), obey all the postulates of common algebra; so do the kinematic operators which we shall consider in this section. It is possible to establish a complete correspondence between the complex numbers of Section 1 and the kinematic operators of this section. Thus we can " apply " complex numbers to certain problems of kinematics. Our intentions, however, are different; we wish to give more substance to the idea of complex numbers. From the point of view of logic this is a step in the wrong direction. If we wish to study a specified set of relations, we should select things that exemplify these relations with as few extraneous properties as possible. Logically geometry and mechanics are more complicated than algebraic expressions of the form $a + bi$. The logician thinks that even the symbol $a + bi$ has more substance than necessary since i is supposed to be a root of (1) and yet in the complex number itself i is nothing more than a separation symbol which keeps a apart from b; therefore, he prefers (a,b) as a symbol for the complex number and likes to reduce the arithmetic of complex numbers to the arithmetic of pairs of real numbers.

On the other hand, what is logically simple is not necessarily intuitively

simple. Many familiar concepts are accepted intuitively long before they are developed on a logical foundation. For this reason, engineers and scientists who are habitually concerned with concrete phenomena find abstract mathematics difficult — "strange" is really the proper word.

In order to develop the conception of complex numbers as kinematic operators, we shall consider *radii*, or "bound" vectors, drawn from a *fixed* point in a plane, Figure 1.1. We shall denote these radii by their end points A, B, C. By definition, the "sum" of two radii, A and B, is the radius P coinciding with the diagonal of the parallelogram constructed on A and B. This definition satisfies the commutative and associative laws.

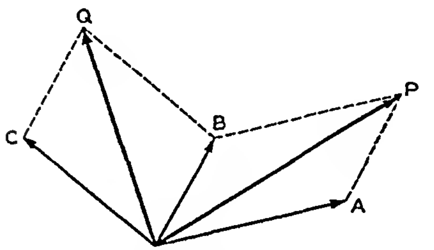

FIG. 1.1. The parallelogram law of addition of radii (bound and directed line segments).

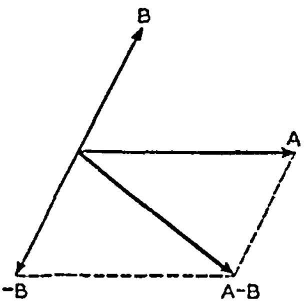

FIG. 1.2. Subtraction of a radius is addition of this radius after it has been rotated through 180°.

The radius of zero length is the "zero" in the algebra of radii. The additive inverse of A is the radius of the same length but in the opposite direction; the sum of the two is zero. Subtraction consists in adding the inverse, Figure 1.2.

The product xA of a radius and a positive real number is another radius in the same direction but of length extended in the ratio x to 1. If x is negative, then xA and A are in opposite directions. Next we define multiplication by "the imaginary unit i" to be a counterclockwise rotation through 90°, Figure 1.3d. Two successive rotations are equivalent to the reversal of direction; thus $ii = i^2 = -1$. Finally, we define the operation of multiplication of a radius by a composite or *complex* number $x + iy$ in accordance with the following equation

$$(x + iy)A = xA + iyA; \tag{10}$$

that is, to form $(x + iy)A$ we form radii xA, iyA and add them.

Addition of complex numbers is defined by

$$[(x + iy) + (u + iv)]A = (x + iy)A + (u + iv)A. \tag{11}$$

Expanding the right side according to (10) we find that it is also the

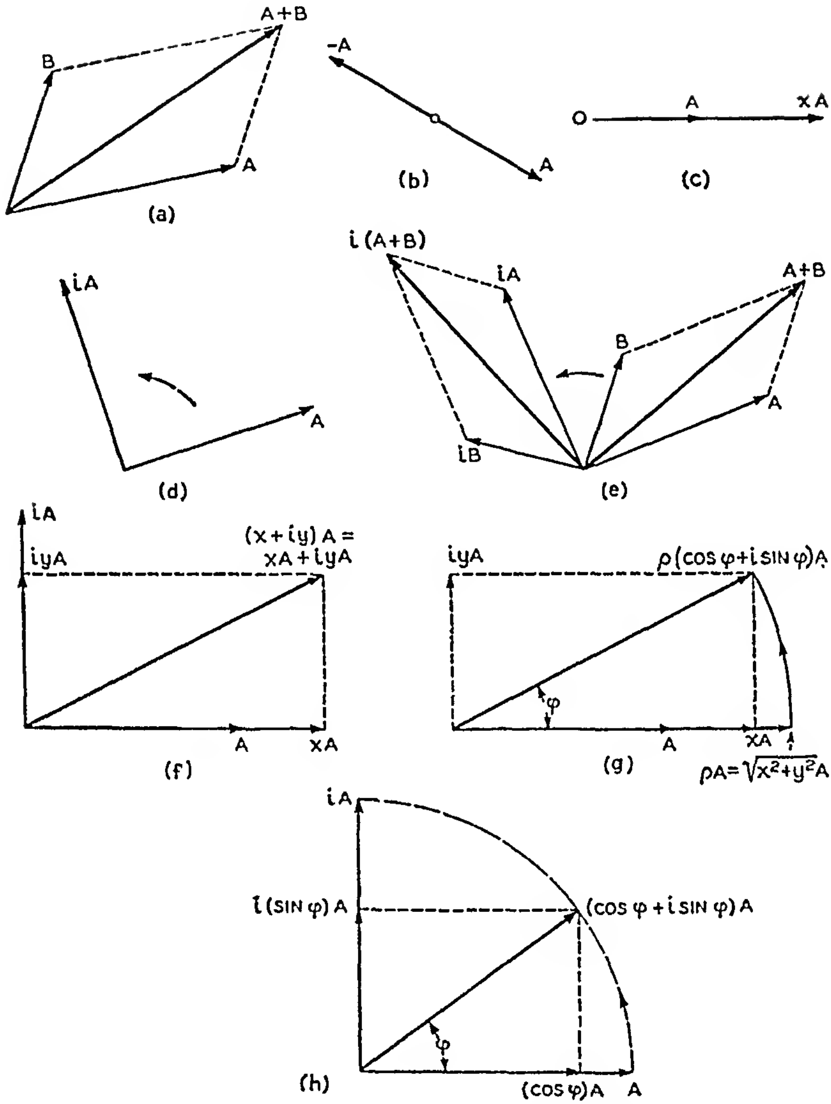

FIG. 1.3. (a) Addition of radii; (b) An additive inverse of a radius; (c) Extension of a radius in the ratio $x/1$, x being positive; if x is negative, xA and A are in opposite directions; (d) Definition of "i" as a counterclockwise rotation through 90°; (e) The same operation as (d); (f) Definition of the complex number $x + iy$ as a kinematic operator on a radius; (g) Illustration of the equivalence of operation (f) to an extension in the ratio $\sqrt{x^2 + y^2}/1$ followed by rotation through an angle φ; (h) Interpretation of unit complex numbers as rotational operators.

expansion of $[(x + u) + i(y + v)]A$. Thus our operators obey the rule of addition (4).

The product of two complex numbers is defined as that complex number which yields the same result when operating on some radius as the successive application of the operations denoted by the factors. Thus $C = (x + iy)(u + iv)A = (x + iy)B$, where $B = (u + iv)A$. We already know how to construct $B = uA + ivA$, $xB = xuA + ixvA$, $iyB = iyuA + i^2yvA = iyuA - yvA$, and finally $C = (x + iy)B = xB + iyB = xuA + ixvA + iyuA - yvA$. Since this radius is derived also from $[(xu - yv) + i(xv + yu)]A$, we obtain the multiplication rule (6).

There is another much simpler multiplication rule which is based on the equivalence of the construction in Figure 1.3f to extension in the ratio $\rho = \sqrt{x^2 + y^2}$ to 1 followed by rotation through angle φ defined by

$$\cos\varphi = x/\rho, \qquad \sin\varphi = y/\rho, \qquad \rho = \sqrt{x^2 + y^2}. \tag{12}$$

This second construction is illustrated in Figure 1.3g. Successive application of this construction is easy since the operations of extension and rotation are commutative; and successive rotations through angles φ_1 and φ_2 are equivalent to a rotation through $\varphi_1 + \varphi_2$. Thus, if

$$x + iy = \rho(\cos\varphi + i\sin\varphi), \tag{13}$$

then

$$\begin{aligned}(x_1 + iy_1)(x_2 + iy_2) &= [\rho_1(\cos\varphi_1 + i\sin\varphi_1)][\rho_2(\cos\varphi_2 + i\sin\varphi_2)] \\ &= \rho_1\rho_2[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)].\end{aligned} \tag{14}$$

Division is an inverse operation and

$$\frac{\rho_1(\cos\varphi_1 + i\sin\varphi_1)}{\rho_2(\cos\varphi_2 + i\sin\varphi_2)} = \frac{\rho_1}{\rho_2}[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)]. \tag{15}$$

Extension is annulled by contraction and counterclockwise rotation by an equal clockwise rotation; thus

$$[\rho(\cos\varphi + i\sin\varphi)]\left\{\frac{1}{\rho}[\cos(-\varphi) + i\sin(-\varphi)]\right\} = 1. \tag{16}$$

Problem. Show in two ways that, if A and B are two radii, $(x + iy)(A + B) = (x + iy)A + (x + iy)B$.

Hint: One method is based on Figure 1.3f and the other on 1.3g; note that $(x + iy)(A + B) = x(A + B) + iy(A + B)$ follows from the definition of multiplication of a radius by $x + iy$ but the equation in the problem needs justification.

5. *Complex plane*

Choosing some particular radius to represent unity, we are enabled to represent every other complex number by one and only one radius or by its

end point in the "complex plane," Figure 1.4. Conversely, points in a plane may be represented by complex numbers.

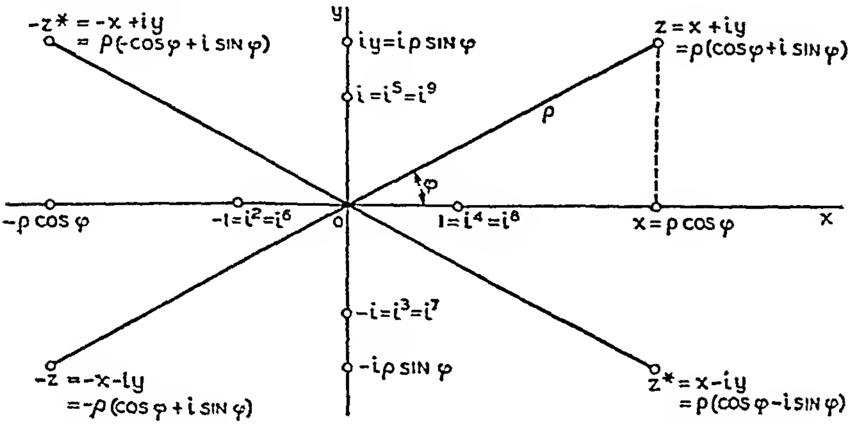

FIG. 1.4. Representation of a complex number by a radius or its end point in the complex plane.

The real and imaginary parts of the complex number z are denoted as follows

$$x = \text{re}\,(z), \qquad y = \text{im}\,(z). \tag{17}$$

The quantity ρ is called the *amplitude*, or *modulus* or *absolute value* of the complex number and is denoted either by am (z) or by $|z|$; thus

$$\text{am}\,(z) = |z| = \rho = \sqrt{x^2 + y^2}. \tag{18}$$

The angle φ defined by

$$\cos\varphi = x/\rho, \qquad \sin\varphi = y/\rho \tag{19}$$

is called the *phase* or *argument* of the complex number and is denoted by

$$\varphi = \text{ph}\,(z). \tag{20}$$

The *conjugate complex number* is defined by

$$z^* = x - iy = \rho(\cos\varphi - i\sin\varphi). \tag{21}$$

Graphically it is represented by the mirror image of z in the x-axis. The rotations implied by z and z^* are equal but in opposite directions; hence,

$$zz^* = \rho^2, \qquad \rho = \sqrt{zz^*}. \tag{22}$$

Problems

1. Show that

a) $\left|\sqrt{x+iy}\right| = \sqrt[4]{x^2+y^2}$,

b) $\left|(x+iy)\sqrt[3]{x+iy}\right| = \sqrt{x^2+y^2}\sqrt[6]{x^2+y^2}$,

c) $\left|\dfrac{x+iy}{u+iv}\right| = \dfrac{\sqrt{x^2+y^2}}{\sqrt{u^2+v^2}}$, d) $\left|\dfrac{\sqrt{x-iy}}{\sqrt[3]{u+iv}}\right| = \dfrac{\sqrt[4]{x^2+y^2}}{\sqrt[6]{u^2+v^2}}$.

2. Show that

$$\left|\frac{x-iy}{x+iy}\right| = 1, \quad \operatorname{ph}\left(\frac{x-iy}{x+iy}\right) = -2\tan^{-1}(y/x).$$

3. Show that

a) $\left|\dfrac{(x+iy)^2}{(x-iy)^2}\right| = 1$, b) $\left|\dfrac{(x+iy)^3}{(x-iy)^2}\right| = \sqrt{x^2+y^2}$.

4. Show that if the point represented by $x+iy$ is inside the circle of unit radius, centered at the origin, then the point $1/(x+iy)$ is outside the circle.

6. *Geometric applications*

Propositions of plane geometry may be established by manipulating complex numbers and interpreting the resulting equations. To illustrate this we shall prove the following proposition: *the line joining the mid-points of two sides of a triangle is parallel to the third side and is equal to one half of it.*

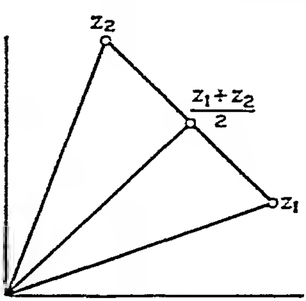

FIG. 1.5. Complex numbers representing the end points and midpoints of a linear segment.

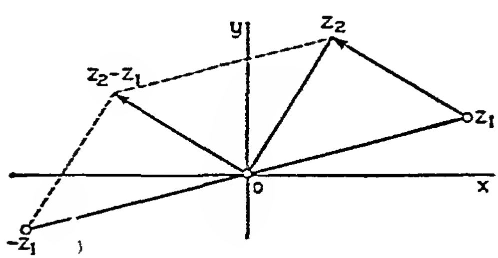

FIG. 1.6 The difference $z_2 - z_1$ represents the magnitude and direction of the segment starting at z_1 and ending at z_2.

First we note that the mid-point of the segment joining two points z_1, z_2 is $(z_1+z_2)/2$, Figure 1.5; next we observe that the line joining z_1 to z_2 is equal and parallel to $z_2 - z_1$, Figure 1.6; the mid-points of two sides of a triangle, Figure 1.7, are $(z_1+z_3)/2$ and $(z_2+z_3)/2$; the line joining them is equal and parallel to $(z_2-z_1)/2$ which is one half of the radius equal and parallel to the third side.

Another proposition is: *the lines joining the mid-points of opposite sides of any quadrilateral bisect each other*, Figure 1.8. The mid-point of EG is $(z_1 + z_2 + z_3 + z_4)/4$; which is also the mid-point of HF.

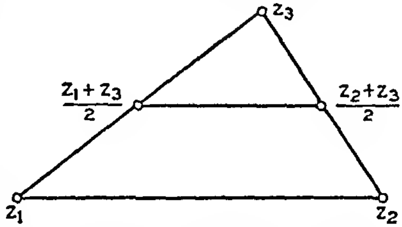

FIG. 1.7. Use of complex numbers in proving a geometric theorem.

Next let us obtain the equation of a straight line joining points z_1 and z_2. If z is any point on this line, Figure 1.9, then the phases of $z - z_1$ and $z - z_2$ are either equal or differ by π; hence the quotient $(z - z_1)/(z - z_2)$ is a *real* number t and the parametric equation of the straight line is

$$z - z_1 = t(z - z_2) \qquad \text{or} \qquad z = (z_1 - tz_2)/(1 - t). \tag{23}$$

Another form of the equation of a straight line is free from parameter t.

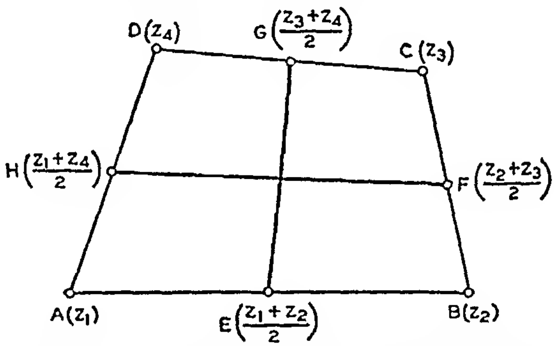

FIG. 1.8. An illustration of a theorem about quadrilaterals.

Since t is real, it is equal to its own conjugate; therefore

$$\frac{z - z_1}{z - z_2} = \frac{z^* - z_1^*}{z^* - z_2^*}, \qquad z(z_1^* - z_2^*) - z^*(z_1 - z_2) + (z_1 z_2^* - z_1^* z_2) = 0. \tag{24}$$

The equation of a circle of radius r concentric with the origin is simply

$$zz^* = r^2. \tag{25}$$

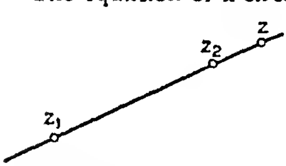

FIG. 1.9. For any point z on a straight line passing through z_1 and z_2 the ratio $(z - z_1)/(z - z_2)$ is real.

The equation of a circle with center at C, Figure 1.10, is

$$(z - C)(z^* - C^*) = r^2. \tag{26}$$

To obtain this equation we need only recall that $z - C$ is the line joining the center of the circle to a point on the circumference, that am $(z - C) = r$, and that the product of conjugate numbers equals the square of the amplitude.

Let us compare (26) with the general cartesian equation of the circle

$$x^2 + y^2 + ax + by + c = 0. \tag{27}$$

In view of

$$x = \tfrac{1}{2}(z + z^*), \qquad iy = \tfrac{1}{2}(z - z^*), \qquad x^2 + y^2 = zz^*, \tag{28}$$

equation (27) becomes

$$zz^* + \tfrac{1}{2}(a - ib)z + \tfrac{1}{2}(a + ib)z^* + c = 0. \tag{29}$$

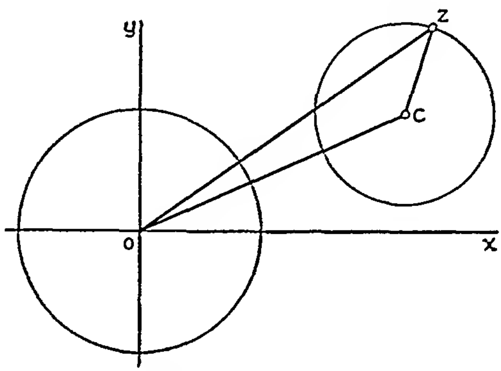

FIG. 1.10. An illustration for the equation of a circle in terms of complex numbers.

Comparing with the expanded form of (26), we obtain

$$C = -\tfrac{1}{2}(a + ib), \qquad r = \sqrt{CC^* - c}. \tag{30}$$

It is evident from (28) that cartesian equations of plane curves may be converted into equations depending on z and z^*. Likewise, $z = f(t) + ig(t)$, where t is a real parameter, represents a curve whose parametric equations are $x = f(t)$, $y = g(t)$.

Problems

1. Show that the mid-point of the line joining the mid-points of the diagonals of a quadrilateral coincides with the point of intersection of the lines joining the mid-points of opposite sides of the quadrilateral.

2. Show that the lines joining the mid-points of the adjacent sides of a quadrilateral form a parallelogram.

3. Show that the cartesian equation $ax + by + c = 0$ of a straight line becomes $Az + A^*z^* + C = 0$, where C is real and $A = a - ib$.

4. Equation (24) is of the form $Mz - M^*z^* + N = 0$. Reconcile this form with that given in the preceding problem.

5. Show that if a and b are two sides of a triangle and φ is the angle between them, then the third side c is given by $c^2 = a^2 - 2ab\cos\varphi + b^2$.

7. Applications to circular functions

Numerous relations between circular functions can readily be obtained with the aid of the algebra of complex numbers. A *unit complex number*

$$U = \cos\varphi + i\sin\varphi \tag{31}$$

represents a rotation through an angle φ. Multiplying two unit numbers we obtain either

$$U_1U_2 = \cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2), \tag{32}$$

or

$$U_1U_2 = (\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + i(\sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2). \tag{33}$$

Equating these results, we have the addition formulas for circular functions

$$\begin{aligned}\cos(\varphi_1 + \varphi_2) &= \cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2,\\ \sin(\varphi_1 + \varphi_2) &= \sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2.\end{aligned} \tag{34}$$

De Moivre's theorem,

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi, \tag{35}$$

is merely a statement of the equivalence of n successive rotations through an angle φ to a single rotation through the angle $n\varphi$. Using the binomial expansion, we can express $\cos n\varphi$ and $\sin n\varphi$ in terms of powers of $\cos\varphi$ and $\sin\varphi$; thus, if $n = 3$,

$$\begin{aligned}&\cos^3\varphi + 3i\cos^2\varphi\sin\varphi + 3i^2\cos\varphi\sin^2\varphi + i^3\sin^3\varphi\\ &\qquad = \cos 3\varphi + i\sin 3\varphi,\\ &\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi,\ \sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi.\end{aligned} \tag{36}$$

Conversely, powers of $\cos\varphi$ and $\sin\varphi$ can be expressed in terms of functions of multiple angles. Noting that

$$\cos\varphi = \frac{1}{2}(U + U^{-1}), \qquad \sin\varphi = \frac{1}{2i}(U - U^{-1}), \tag{37}$$

$$\cos n\varphi = \frac{1}{2}(U^n + U^{-n}), \qquad \sin n\varphi = \frac{1}{2i}(U^n - U^{-n}), \tag{38}$$

we raise (37) to the nth power

$$\cos^n\varphi = \frac{1}{2^n}(U + U^{-1})^n, \qquad \sin^n\varphi = \frac{1}{2^n i^n}(U - U^{-1})^n. \tag{39}$$

Applying the binomial theorem and reducing with the aid of (38), we ob-

tain the desired expressions. For example, if $n = 2$, then

$$\begin{aligned} \cos^2 \varphi &= \tfrac{1}{4}(U^2 + 2 + U^{-2}) = \tfrac{1}{2}(\cos 2\varphi + 1), \\ \sin^2 \varphi &= -\tfrac{1}{4}(U^2 - 2 + U^{-2}) = \tfrac{1}{2}(1 - \cos 2\varphi). \end{aligned} \tag{40}$$

De Moivre's theorem yields solutions of the following equation,

$$z^n = 1, \tag{41}$$

where n is an integer. If $n\varphi = 2\pi$, the number on the right in (35) is unity; therefore,

$$z_1 = \cos (2\pi/n) + i \sin (2\pi/n) \tag{42}$$

is a root of (41). If the nth power of z_1 is unity, the nth power of the square of z_1 or of any other integral power of z_1 is also unity; hence all the roots may be expressed as follows

$$z_1, z_1^2, z_1^3, \cdots z_1^{n-1}, z_1^n = 1, \qquad z_1^m = \cos (2m\pi/n) + i \sin (2m\pi/n). \tag{43}$$

The $(n + 1)$th power of z_1 equals z_1 and the sequence repeats itself.

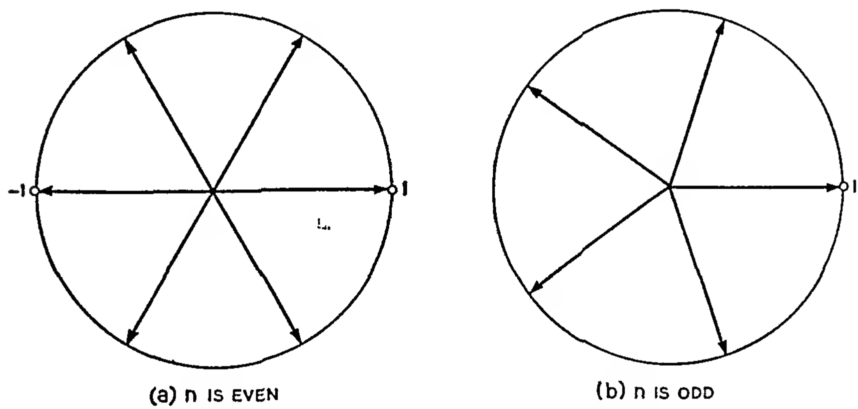

FIG. 1.11. A visual proof that the sum of the roots of $z^n = 1$ is 0.

These roots are represented by equispaced radii of the unit circle, Figure 1.11. By a theorem of algebra the sum of the roots of the algebraic equation, $z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0$, equals $-a_{n-1}$; in (41) this coefficient is zero and therefore

$$1 + z_1 + z_1^2 + \cdots + z_1^{n-1} = 0. \tag{44}$$

This result may also be obtained directly from Figure 1.11. The rules for addition of complex numbers and for addition of forces are the same; the resultant of symmetrically disposed equal forces must vanish or else we should be at a loss to prescribe it a definite direction; hence, conclusion

(44). Equating the real and imaginary parts, we have

$$
\begin{aligned}
1 + \cos(2\pi/n) + \cos(4\pi/n) + \cdots + \cos(2\overline{n-1}\pi/n) &= 0, \\
\sin(2\pi/n) + \sin(4\pi/n) + \cdots + \sin(2\overline{n-1}\pi/n) &= 0.
\end{aligned} \tag{45}
$$

In the theory of radiation from a linear array of radiators of electromagnetic or sound waves the following series are encountered:

$$
\begin{aligned}
P &= 1 + \cos\varphi + \cos 2\varphi + \cdots + \cos n\varphi, \\
Q &= \sin\varphi + \sin 2\varphi + \cdots + \sin n\varphi.
\end{aligned} \tag{46}
$$

To obtain the sums, we form a third series

$$
\begin{aligned}
R = P + iQ = 1 &+ (\cos\varphi + i\sin\varphi) + (\cos 2\varphi + i\sin 2\varphi) \\
&+ \cdots + (\cos n\varphi + i\sin n\varphi),
\end{aligned} \tag{47}
$$

which is the geometric series

$$R = 1 + U + U^2 + \cdots + U^n. \tag{48}$$

Multiplying by U and subtracting the result from (48), we get

$$R(1 - U) = 1 - U^{n+1}, \qquad R = \frac{1 - U^{n+1}}{1 - U}. \tag{49}$$

The real and imaginary parts may be separated as follows

$$
\begin{aligned}
R &= \frac{U^{n+1} - 1}{U - 1} = \frac{U^{(n+1)/2}[U^{(n+1)/2} - U^{-(n+1)/2}]}{U^{1/2}[U^{1/2} - U^{-1/2}]} \\
&= [\cos(n\varphi/2) + i\sin(n\varphi/2)]\frac{\sin[(n+1)\varphi/2]}{\sin(\varphi/2)}, \qquad (50) \\
P &= \frac{\cos(n\varphi/2)\sin[(n+1)\varphi/2]}{\sin(\varphi/2)}, \\
Q &= \frac{\sin(n\varphi/2)\sin[(n+1)\varphi/2]}{\sin(\varphi/2)}, \\
|R| &= \left|\frac{\sin[(n+1)\varphi/2]}{\sin(\varphi/2)}\right|.
\end{aligned}
$$

Summation of many infinite series may be effected by a similar method.

Problems

1. Use De Moivre's theorem to obtain

$$
\begin{aligned}
&\cos 2\varphi = \cos^2\varphi - \sin^2\varphi, \ \sin 2\varphi = 2\sin\varphi\cos\varphi, \\
&\cos 4\varphi = \cos^4\varphi - 6\cos^2\varphi\sin^2\varphi + \sin^4\varphi, \\
&\sin 4\varphi = 4\cos^3\varphi\sin\varphi - 4\cos\varphi\sin^3\varphi.
\end{aligned}
$$

2. Show that

$$\cos n\varphi = \cos^n \varphi - \frac{n(n-1)}{2!}\cos^{n-2}\varphi \sin^2\varphi + \frac{n(n-1)(n-2)(n-3)}{4!}\cos^{n-4}\varphi\sin^4\varphi - \cdots$$

$$\sin n\varphi = n \sin\varphi \cos^{n-1}\varphi - \frac{n(n-1)(n-2)}{3!}\sin^3\varphi\cos^{n-3}\varphi + \cdots.$$

3. Derive

$$\cos^n\varphi = 2^{-n+1}\left[\cos n\varphi + n\cos(n-2)\varphi + \frac{n(n-1)}{2!}\cos(n-4)\varphi + \cdots\right],$$

where the last term is $n(n-1)\cdots[(n+3)/2](\cos\varphi)/[(n-1)/2]!$ if n is odd, and $n(n-1)\cdots[(n/2)+1]/(n/2)!\,2$ if n is even.

4. Show that if n is even,

$$\sin^n\varphi = (-)^{n/2}2^{-n+1}\left[\cos n\varphi - n\cos(n-2)\varphi + \frac{n(n-1)}{2!}\cos(n-4)\varphi + \cdots\right];$$

if n is odd,

$$\sin^n\varphi = (-)^{(n-1)/2}2^{-n+1}\left[\sin n\varphi - n\sin(n-2)\varphi + \frac{n(n-1)}{2!}\sin(n-4)\varphi - \cdots\right].$$

5. Show that

$$\begin{aligned}\cos(A+B+C) &= \cos A\cos B\cos C - \sin A\sin B\cos C\\ &\quad - \sin A\cos B\sin C - \cos A\sin B\sin C,\\ \sin(A+B+C) &= \sin A\cos B\cos C + \cos A\sin B\cos C\\ &\quad + \cos A\cos B\sin C - \sin A\sin B\sin C.\end{aligned}$$

6. Find the roots of $z^n = -1$.

7. Find the roots of

$$z^n = r(\cos\theta + i\sin\theta).$$

8. Show that

$$\cos\varphi + \cos\left(\varphi + \frac{2\pi}{n}\right) + \cos\left(\varphi + \frac{4\pi}{n}\right) + \cdots + \cos\left[\varphi + \frac{2(n-1)\pi}{n}\right] = 0,$$

$$\sin\varphi + \sin\left(\varphi + \frac{2\pi}{n}\right) + \sin\left(\varphi + \frac{4\pi}{n}\right) + \cdots + \sin\left[\varphi + \frac{2(n-1)\pi}{n}\right] = 0.$$

Hint: It is possible to conclude that these equations are true by just looking at a certain picture.

9. Show that

$$1 + 4\rho\cos\varphi + 6\rho^2\cos 2\varphi + 4\rho^3\cos 3\varphi + \rho^4\cos 4\varphi = (1 + 2\rho\cos\varphi + \rho^2)^2\cos\left[4\tan^{-1}\frac{\rho\sin\varphi}{1+\rho\cos\varphi}\right].$$

8. *Transformation of potential functions*

Complex variables are useful in effecting a change of coordinates in two-dimensional and in some three-dimensional potential problems. To illustrate, we shall take a potential function*

$$V_n = \rho^n \cos n\varphi \tag{51}$$

in polar coordinates and transform it into another polar system obtained by translation of the axis, Figure 1.12. Direct transformation would be long. We may, however, take advantage of the fact that (51) is the real part of

$$W_n = z^n,$$
$$z = \rho(\cos\varphi + i\sin\varphi). \tag{52}$$

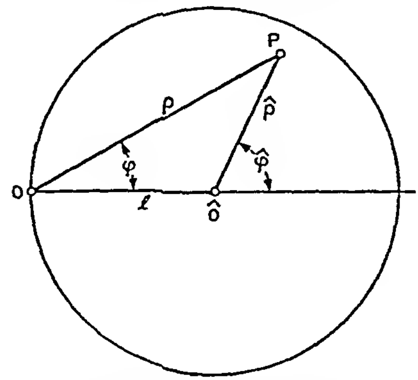

FIG. 1.12. Transformation of potential functions from one coordinate system to another is facilitated by using complex variables.

This new function is called the *complex potential.* Introducing

$$\hat{z} = \hat{\rho}(\cos\hat{\varphi} + i\sin\hat{\varphi}), \tag{53}$$

we have, from Figure 1.12,

$$z = \ell + \hat{z}; \tag{54}$$

therefore,

$$W_n = (\ell + \hat{z})^n. \tag{55}$$

If n is a positive integer, the elementary binomial theorem yields

$$W_n = \ell^n + n\ell^{n-1}\hat{z} + \frac{n(n-1)}{1\cdot 2}\ell^{n-2}\hat{z}^2 + \cdots + \hat{z}^n. \tag{56}$$

Taking the real part, we have

$$V_n = \ell^n + n\ell^{n-1}\hat{\rho}\cos\hat{\varphi} + \frac{n(n-1)}{1\cdot 2}\ell^{n-2}\hat{\rho}^2\cos 2\hat{\varphi} + \cdots + \hat{\rho}^n\cos n\hat{\varphi}. \tag{57}$$

If n is a negative integer, the transformation leads to an infinite series; for in this case the binomial expansion is

$$(1+z)^n = 1 + nz + \frac{n(n-1)}{1\cdot 2}z^2 + \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}z^3 + \cdots, \ |z| < 1. \tag{58}$$

The series is convergent only when the amplitude of z is less than unity.

* Any solution of Laplace's equation (3), Chapter 13; but the precise meaning of the term "potential function" is not important for what follows.

Hence we have to consider two cases: (1) the amplitude of $\hat{z}$ is less than ℓ and P is inside the circle with center at $\hat{O}$ and passing through O, Figure 1.12; (2) $\operatorname{am}(\hat{z}) > \ell$ and P is outside this circle. In the two cases we have

$$
\begin{aligned}
W_n &= \ell^n\left(1+\frac{\hat{z}}{\ell}\right)^n = \ell^n\left[1+n\frac{\hat{z}}{\ell}+\frac{n(n-1)}{1\cdot 2}\left(\frac{\hat{z}}{\ell}\right)^2+\cdots\right],\\
W_n &= \hat{z}^n\left(1+\frac{\ell}{\hat{z}}\right)^n = \hat{z}^n\left[1+n\frac{\ell}{\hat{z}}+\frac{n(n-1)}{1\cdot 2}\left(\frac{\ell}{\hat{z}}\right)^2+\cdots\right].
\end{aligned}
\tag{59}
$$

Taking the real parts, we have

$$
\begin{aligned}
V_n &= \ell^n\left[1+n\left(\frac{\hat{\rho}}{\ell}\right)\cos\hat{\varphi}+\frac{n(n-1)}{1\cdot 2}\left(\frac{\hat{\rho}}{\ell}\right)^2\cos 2\hat{\varphi}+\cdots\right]\\
&= \hat{\rho}^n\cos n\hat{\varphi}+n\hat{\rho}^n\left(\frac{\ell}{\hat{\rho}}\right)\cos(n-1)\hat{\varphi}\\
&\quad+\frac{n(n-1)}{1\cdot 2}\hat{\rho}^n\left(\frac{\ell}{\hat{\rho}}\right)^2\cos(n-2)\hat{\varphi}+\cdots.
\end{aligned}
\tag{60}
$$

The reason why this method of transforming coordinates in potential functions turns out to be successful will be understood when we come to study the partial differential equation for the potentials and exhibit its relation to functions of a complex variable. If we were to try to apply the method to such functions as $\rho^3 \cos 2\varphi$, we would get nowhere. Most functions are not real parts of functions obtained by arithmetic operations on a complex variable; but the two-dimensional potential functions always are.

Problems

1. Transform $V_n = \rho^n \sin n\varphi$ into the polar system with the origin at $(\ell, 0)$, assuming that the polar axes coincide.

2. Express (51) in cartesian coordinates.

Ans. $$V_n = x^n - \frac{n(n-1)}{2!}x^{n-2}y^2 + \frac{n(n-1)(n-2)(n-3)}{4!}x^{n-4}y^4 - \cdots$$

if n is a positive integer; if n is a negative integer, $-m$, then

$$V_n = (x^2+y^2)^{-m}\left[x^m - \frac{m(m-1)}{2!}x^{m-2}y^2+\cdots\right].$$

3. Find general transformations for (51) from one polar system into another (arbitrary origins and directions of the axes).

9. *Differentiation of unit complex variables*

Let the amplitude of the complex variable be unity as in (31). Differentiating with respect to φ, we have

$$dU/d\varphi = -\sin\varphi + i\cos\varphi. \tag{61}$$

The derivative of U can be obtained directly from first principles. Let φ take on an increment $\Delta\varphi$, Figure 1.13. The corresponding increment ΔU is represented by the directed segment drawn from A to B. Since $\Delta\varphi$ is real, the average derivative $\Delta U/\Delta\varphi$ has the same direction as AB. The absolute value of this ratio approaches unity as $\Delta\varphi$ decreases for the ratio of the lengths of the arc and chord tends to unity. In the limit $\Delta U/\Delta\varphi$ becomes perpendicular to U. To summarize: $dU/d\varphi$ is U turned counterclockwise through 90°

$$dU/d\varphi = iU = i\,(\cos\varphi + i\sin\varphi). \tag{62}$$

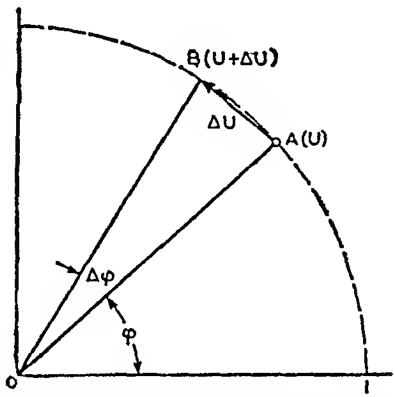

FIG. 1.13. The derivative of a unit complex variable with respect to its phase is equal to the variable multiplied by i.

Having obtained this equation, we can deduce the derivatives of $\cos\varphi$ and $\sin\varphi$ by equating the real and imaginary parts

$$\frac{d}{d\varphi}(\cos\varphi + i\sin\varphi) = i\cos\varphi - \sin\varphi,$$

$$\frac{d}{d\varphi}(\cos\varphi) = -\sin\varphi, \qquad \frac{d}{d\varphi}(\sin\varphi) = \cos\varphi. \tag{63}$$

The operational interpretation of (62) is worth keeping constantly in mind. In substance, there is no difference between the above method for differentiating U and the usual method for differentiating circular functions; and yet, the combination $\cos\varphi + i\sin\varphi$ has simpler properties than its components.

10. *Harmonic oscillations*

As kinematic operators, complex numbers are well adapted to analytic representation of rotary motion and subsequently of harmonic oscillations. This application started on a very elementary plane but has gradually developed into one of the most important methods of applied mathematics.

Consider a radius OP, Figure 1.14, revolving about O with the angular velocity ω radians per second, that is, with the frequency $f = \omega/2\pi$ cycles per second. The angular position φ of the radius is called the *phase* of the motion; in the present case it is a linear function of time

$$\varphi = \omega t + \varphi_0, \tag{64}$$

where φ_0 is the *initial phase*. If the length of the radius is a, the motion of the end point may be expressed by a complex variable

$$z = a(\cos\varphi + i\sin\varphi) = a[\cos(\omega t + \varphi_0) + i\sin(\omega t + \varphi_0)]. \tag{65}$$

The real and imaginary parts of z are also harmonic functions of time. Between these three functions there is a unique correspondence; if we

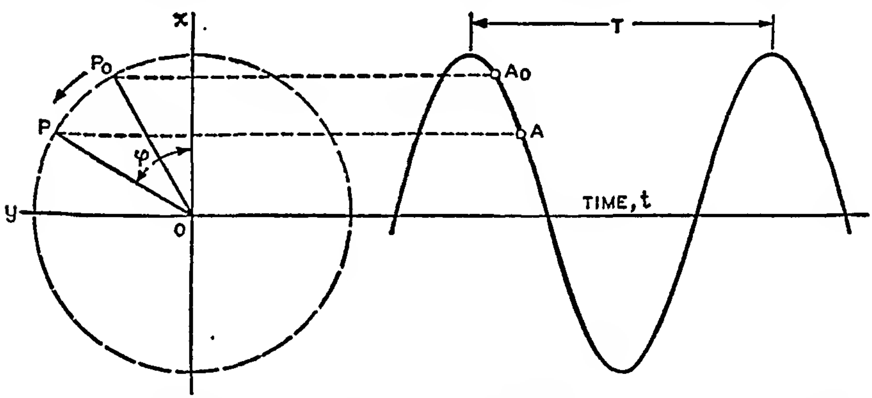

FIG. 1.14. Harmonic motion of a given frequency can be correlated with uniform circular motion, and both may be represented by complex quantities of constant magnitude but with phase proportional to time.

know one, we know the others. Figure 1.14 depicts specifically the correspondence between z and its real part

$$x = \mathrm{re}\,(z) = a\cos\varphi = a\cos(\omega t + \varphi_0). \tag{66}$$

Time T required for one complete revolution or oscillation is the *period*. Evidently

$$fT = 1, \qquad \omega T = 2\pi. \tag{67}$$

In dealing with several harmonic variables of the *same frequency* we can confine our attention solely to their amplitudes and initial phases. Introducing the *complex amplitude*

$$A = a(\cos\varphi_0 + i\sin\varphi_0), \tag{68}$$

and the *time factor*

$$\mathfrak{Z} = \cos\omega t + i\sin\omega t, \tag{69}$$

we change (65) into $$z = A\mathfrak{Z}. \tag{70}$$

From the kinematic point of view this equation is self-evident; A is the position of the radius at $t = 0$ and $\mathfrak{Z}$ rotates it through an angle directly proportional to time.

Suppose we wish to add several harmonic variables of the same frequency. We may want the total voltage across a resistor, capacitor, and inductor in series; or the total electric current through several circuit elements in parallel; or the total excess pressure due to several sound waves arriving from different points. Since

$$A\mathfrak{Z} + B\mathfrak{Z} + C\mathfrak{Z} = (A + B + C)\mathfrak{Z}, \tag{71}$$

our problem is reduced to the addition of the complex amplitudes of the component oscillations. The operation can be performed graphically as in Figure 1.15.

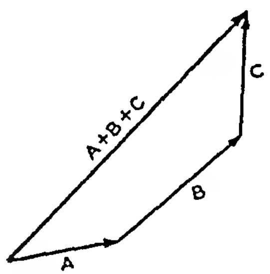

FIG. 1.15. A graphical construction for obtaining the amplitude and phase of several sinusoids of the same frequency.

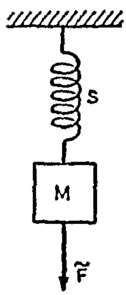

FIG. 1.16. A force $\tilde{F}$ acting on a mass M attached to a spring with stiffness S.

This simple application is just a beginning. Complex variables can be used to solve certain types of differential equation with no more effort than that involved in multiplication and division. Let us calculate the steady state motion of a mass M attached to a spring of stiffness S and acted upon by a harmonic force, Figure 1.16. Let

$\tilde{F} = F_a \cos(\omega t + \varphi_F)$ be the instantaneous value of the applied force,
$F_a =$ the amplitude,
$\varphi_F =$ the initial phase,
$F = F_a(\cos\varphi_F + i \sin\varphi_F) =$ the complex amplitude,
$\breve{F} = F\mathfrak{Z}$ so that
$\tilde{F} = \operatorname{re}\breve{F} = \operatorname{re} F\mathfrak{Z}$.

We shall use a similar set of symbols for the displacement of the mass from its neutral position.

The equation of motion is

$$M\frac{d^2\tilde{y}}{dt^2} + R\frac{d\tilde{y}}{dt} + S\tilde{y} = \tilde{F}, \tag{72}$$

where R is the resistance coefficient. This is the equation connecting any displacement of the mass to the action of the given force, in particular the *steady state* displacement $\tilde{y}$ which is *of the same frequency as* $\tilde{F}$. If

$$M\frac{d^2\mathring{y}}{dt^2} + R\frac{d\mathring{y}}{dt} + S\mathring{y} = \mathring{F} \tag{73}$$

is true, (72) is also true; for the equality of two complex numbers implies the equality of their real parts. The time factor is the only function depending on t; hence

$$My\frac{d^2\mathfrak{z}}{dt^2} + Ry\frac{d\mathfrak{z}}{dt} + Sy\mathfrak{z} = F\mathfrak{z}. \tag{74}$$

Since

$$\frac{d\mathfrak{z}}{dt} = i\mathfrak{z}\frac{d(\omega t)}{dt} = i\omega\mathfrak{z}, \tag{75}$$

we have

$$\frac{d^2\mathfrak{z}}{dt^2} = (i\omega)^2\mathfrak{z} = -\omega^2\mathfrak{z}\,, \cdots \frac{d^n\mathfrak{z}}{dt^n} = (i\omega)^n\mathfrak{z}. \tag{76}$$

Substituting in (74), we obtain

$$-\omega^2My + i\omega Ry + Sy = F, \qquad y = \frac{F}{(S-\omega^2M) + i\omega R}. \tag{77}$$

Calculating the amplitude and phase of y, we have

$$y_a = \frac{F_a}{\sqrt{(S-\omega^2M)^2 + \omega^2R^2}}, \qquad \varphi_y = \varphi_F - \tan^{-1}\frac{\omega R}{(S-\omega^2M)}. \tag{78}$$

From this we obtain the instantaneous value $\tilde{y} = y_a \cos(\omega t + \varphi_y)$.

We can obtain $\tilde{y}$ also as follows. We can always assume that some one initial phase is equal to zero because the origin of time is at our disposal. Assuming that the initial phase of $\tilde{F}$ is zero, we have

$$\mathring{y} = y\mathfrak{z} = \frac{F_a(\cos\omega t + i\sin\omega t)}{(S-\omega^2M) + i\omega R}. \tag{79}$$

Multiplying both terms of this fraction by the conjugate of the denominator, we get

$$\mathring{y} = \frac{F_a(\cos\omega t + i\sin\omega t)[(S-\omega^2M) - i\omega R]}{(S-\omega^2M)^2 + \omega^2R^2} = \frac{F_a}{(S-\omega^2M)^2 + \omega^2R^2}$$
$$\times [(S-\omega^2M)\cos\omega t + \omega R\sin\omega t + i(S-\omega^2M)\sin\omega t - i\omega R\cos\omega t].$$

Finally

$$\tilde{y} = \mathrm{re}\,\mathring{y} = \frac{F_a[(S - \omega^2 M)\cos\omega t + \omega R \sin\omega t]}{(S - \omega^2 M)^2 + \omega^2 R^2}. \tag{80}$$

Once the separate steps are understood, the transition from the original equation of motion (72) to the final equation (77) yielding the complex amplitude of the unknown harmonic variable may be accomplished in accordance with the following simple rules: (1) replace each differentiation with respect to t by $i\omega$, (2) replace the instantaneous values of the known and unknown variables by their complex amplitudes. The rest of the problem is reduced to ordinary algebra. The solution obtained in this manner is the steady state solution, for we have assumed *a priori* that the displacement is a harmonic variable of the same frequency as the force. Later the method will be extended to include the transient terms which *are not* of the same frequency as the impressed force.

The method requires that all harmonic variables occur only in the first degree, that there are no cross-products, and that the coefficients in the differential equation are constants. If S were a function of t, the final equation (77) could not be satisfied since every other quantity is constant. If the equation contains the square of $\tilde{y}$, we cannot replace it by the square of $\mathring{y}$ and assert that the real part of the solution of the transformed equation is the solution of the original equation since

$$\mathrm{re}\,(\mathring{y}^2) \neq [\mathrm{re}\,(\mathring{y})]^2. \tag{81}$$

The same argument applies to cross-products of harmonic variables.

Physical systems satisfying the aforementioned requirements are called *invariable, linear systems*. They are invariable because mass, stiffness, inductance, etc., do not vary with time; they are linear because the equations of motion are first degree polynomials in the variables.

The real part of (73) is our original equation (72); what can we say about the imaginary part? We can, of course, ignore it as of no interest. However, the imaginary part is an equation equivalent to (72); if t is replaced by $t + (\pi/2\omega)$ the sines become the cosines.

The ratio y/F of the complex amplitudes is determined solely by the mechanical system and the frequency. Such ratios play a very important role in the theory of oscillations of any kind and some are given special names. Thus the force/velocity ratio is called the *mechanical impedance* of the system; the voltage/current ratio is the *electrical impedance*; the voltage/velocity and force/current ratios are *electromechanical impedances*; etc. The reciprocal of the impedance is the *admittance*. The real and imaginary parts of the impedance are respectively the *resistance* and the *reactance*; the real and imaginary parts of the admittance are the *conductance* and *susceptance*.

The reader can readily verify that the impedance of the mass M is $i\omega M$; the impedance of the spring of stiffness S is $S/i\omega$. Addition of impedances and admittances can obviously be performed graphically. The importance of such graphical methods can hardly be overestimated.

Eventually we shall find that impedance functions can profitably be regarded as functions of a *complex oscillation constant* p of which $i\omega$ is the imaginary part. In this way the theory of functions of a complex variable is introduced in the theory of oscillations and waves and becomes a method of analysis of inestimable practical value.

Problems

1. Find steady state solutions of the following differential equations

a) $2\bar{y}' + \bar{y} = 3\cos\omega t$,

b) $\bar{y}'' + 4\bar{y}' + 25\bar{y} = 6\sin\omega t = 6\cos(\omega t - \frac{1}{2}\pi)$,

c) $\frac{1}{2}\bar{y}''' - \bar{y}'' + 2\bar{y}' - \frac{1}{4}\bar{y} = \cos\omega t + 2\sin\omega t$.

Note: In the last equation convert the right side into the standard form $A\cos(\omega t + \varphi)$. Verify the answers by substitution.

2. Find the steady state solutions of the following systems of equations

a) $\bar{x}' = -\bar{y} + 2\cos\omega t, \qquad \bar{y}' = \bar{x}$;

b) $\bar{x}'' + 3\bar{x}' + 2\bar{y} = \cos\omega t, \qquad \bar{x}' - \bar{y}'' = \sin\omega t$.

Note: The complex amplitude of $\sin\omega t$ is $-i$.

3. Find the steady state solutions of

a) $\bar{x}' + \bar{x} = \cos t + 4\cos 3t$,

b) $\bar{x}'' + 2\bar{x}' + 3\bar{x} = 5\sin 2t + 2\cos 3t$.

Hint: Observe that the sum $\bar{x}_1 + \bar{x}_2$ of the steady state solutions of $\bar{x}_1' + \bar{x}_1 = \cos t$ and $\bar{x}_2' + \bar{x}_2 = 4\cos 3t$ satisfies (a).

11. *Bilinear transformations*

A functional relationship, $w = f(z)$, between two complex variables transforms geometric figures in one plane into corresponding figures in the other. This transformation is the best available method for the graphic portrayal of functions of a complex variable. Since we have to deal with four variables, to represent them as coordinates of a single point would require a space of four dimensions — no help at all as far as visualization is concerned.

Let

$$w = az + b, \tag{82}$$

where a and b are complex constants. If $a = 1$, this transformation represents a displacement of the entire z plane. If $b = 0$, the transformation consists of stretching the plane in all directions from the origin in the ratio $\mathrm{am}(a)/1$, followed by a counterclockwise rotation through the angle $\mathrm{ph}(a)$. The general *linear transformation* (82) can be visualized as stretching and rotation (given by a) followed by translation (given by b); or else as a translation, given by b/a, and a subsequent stretching and rotation.

Next in complexity is the *bilinear* transformation

$$w = \frac{az + b}{cz + d}, \qquad ad - bc \neq 0. \tag{83}$$

If $ad - bc = 0$, then $b/a = d/c$ and w reduces to a constant; the entire z-plane would be transformed into a single point — not an interesting case. Bilinear transformations have long occupied an important place in the theory of functions; but more recently they have become very prominent in microwave transmission theory.

If a transmission line of length ℓ is terminated into an "output impedance" Z, then the "input impedance" is

$$Z_i = K\frac{Z \cos \beta\ell + iK \sin \beta\ell}{K \cos \beta\ell + iZ \sin \beta\ell} = K\frac{Z + iK \tan \beta\ell}{K + iZ \tan \beta\ell}, \tag{84}$$

where K is the "characteristic impedance," $\beta = 2\pi/\lambda$ and λ is the wave length. For most purposes K is regarded as real; but Z and Z_i are complex. For physical reasons the real parts of both impedances are non-negative and the variables are confined to the right half-planes. The relationship between the input and output impedances is seen to be bilinear.

The "apparent reflection coefficient" w at distance ℓ from the output impedance is

$$w = \frac{1 - z_i}{1 + z_i}, \qquad z_i = Z_i/K, \tag{85}$$

This is another bilinear transformation.

The inverse of a bilinear transformation is also a bilinear transformation

$$z = \frac{-dw + b}{cw - a}. \tag{86}$$

The coefficients b, c have kept their places; but a, d have changed their places and reversed their signs.

We shall now prove that *bilinear transformations change circles into circles*. In this connection straight lines are regarded as circles of infinite radius with centers at infinity. First we shall derive a parametric equation

of the circle passing through three given points $A(z_1)$, $B(z_2)$ and $C(z_3)$, Figure 1.17. Let P be some point on the circle. The angles APC and ABC are either equal or supplementary; therefore, the phases of $(z - z_1)/(z - z_3)$ and $(z_2 - z_1)/(z_2 - z_3)$ are either equal* or differ by π. In either case the quotient of these two ratios is a real number t, positive or negative,

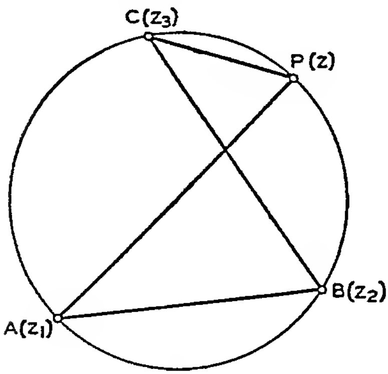

FIG. 1.17. Any point z on a circle may be represented as a function of three assigned points on the circle and one real parameter.

$$\frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)} = t. \tag{87}$$

Solving for z, we have

$$z = \frac{tz_3(z_1 - z_2) + z_1(z_2 - z_3)}{t(z_1 - z_2) + (z_2 - z_3)}. \tag{88}$$

This shows that

$$z = \frac{mt + n}{pt + q}, \tag{89}$$

where t is real, represents a circle.

Suppose now that z in (83) is on a circle. From (87) we can find t to represent each point on this circle; substituting from (88) in (83), we obtain w as a bilinear function of t. As we have already shown, this must represent a circle.

Since $\tan \beta l$ is a real quantity, varying from $-\infty$ to $+\infty$ as l varies, Z_i as given by (84) describes a circle — the so-called "impedance circle." This circle passes through Z when $l = 0$ or more generally when $\beta l = n\pi$, $n = 0, 1, 2, \cdots$; it passes through K^2/Z when $\beta l = (2n + 1)\pi/2$. The circle also passes through Z^*. To prove this we need only show that a real value of l may be found from (84) when $Z_i = Z^*$. Solving (84) for $\tan \beta l$, we find

$$\tan \beta l = K \frac{Z - Z_i}{i(ZZ_i - K^2)}. \tag{90}$$

Since ZZ^* is real and $Z - Z^*$ is imaginary, $\tan \beta l$ is indeed real if $Z_i = Z^*$.

The center of the impedance circle is on the real axis. To prove this, we note: (1) the center of any circle passing through a pair of points lies on the perpendicular bisector of the line joining these points, (2) the perpendicular bisector of the line joining the conjugate points Z, Z^* is the real axis, and (3) Z and Z^* lie on the impedance circle.

* Remembering that the phase of the quotient of two complex numbers is the difference of the angles made by the corresponding radii with the real axis; that is, the angle between the radii.

There is no loss of generality in the assumption that $K = 1$; K merely stretches the impedance plane uniformly in all radial directions from the origin. Henceforward we make this assumption. It has already been stated that, on physical grounds, the real part of $Z = R + iX$ is never negative. If $Z = 1$, then $Z_l = 1$ for all l and the impedance circle reduces to a point. The transmission line is said to be matched and the apparent reflection coefficient is zero at all distances from the output impedance. If $Z = R < 1$, the second point of intersection (in addition to R) of the impedance circle and the real axis is $1/R$; thus one point is inside and the other outside the unit circle. As $R \to 0$, $1/R \to \infty$; the impedance circle will gradually expand and as R varies from unity to zero, will pass by all points in the right half-plane. No two circles intersect each other and the family of impedance circles looks as in Figure 1.18.

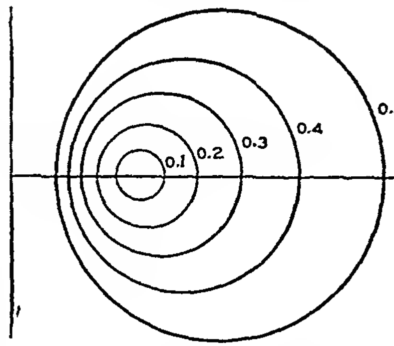

FIG. 1.18. The system of impedance circles for non-dissipative transmission lines terminated into various impedances. Each circle arises from a continuous variation in the length of the line.

From the general theorem about bilinear transformations we conclude that the family of circles (84) when transformed into the "reflection plane" by (85) will remain a family of circles. Substituting from (84) into (85), we have (for $Z = R$)

$$w = \frac{1-R}{1+R}(\cos 2\beta l - i \sin 2\beta l). \tag{91}$$

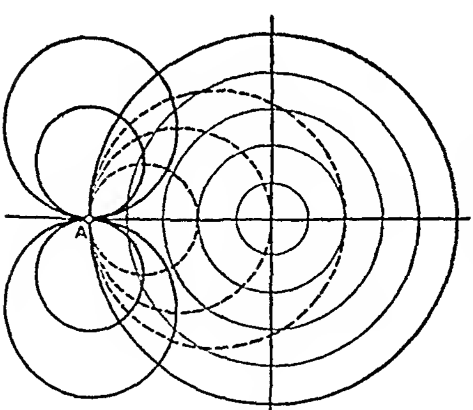

FIG. 1.19. Transformation of the right half of the impedance plane of Fig. 1.18 into the region bounded by a circle is effected by the bilinear transformation (1–84). The impedance circles of the former plane become a system of concentric circles. The new plane is the plane of the reflection coefficient.

These circles are concentric, all inside the unit circle. The latter corresponds to $R = 0, \infty$. The entire physical part of the impedance plane is now inside the unit circle, Figure 1.19.

The lines of constant resistance and variable reactance, being straight lines in the impedance plane, will become circles in the reflection coefficient

plane. All the original straight lines have a point at infinity in common; this point becomes $w = -1$ in the new plane and through this point all circles of constant resistance will pass. When $X = 0$, $w = (1 - R)/(1 + R)$; this gives another point for each particular circle. Since the conjugate values $R + iX$, $R - iX$ are on the same straight line in the Z-plane, and the conjugate values of w correspond to them, the constant resistance circles have their centers on the real axis. The lines of constant reactance X are circles tangential to the real axis at $w = -1$.

Such "impedance charts" are very helpful in the analysis of transmission phenomena. Some charts are based on hyperbolic functions of the complex variable and are not quite as simple in their geometric aspects. A more detailed study of transformations and other examples of their applications will be found in Chapter 14.

Problems

1. Show by direct substitution in equation (29) that the transformation $z = 1/w$ changes circles into circles. This transformation is called the *inversion.*

2. Show by direct substitution in (29) that the transformations $z = Aw$ and $z = w + B$, where A and B are either real or complex, change circles into circles.

3. Show by direct substitution in (29) that the general bilinear transformation $z = (Aw + B)/(Cw + D)$ changes circles into circles.

4. Show that the bilinear transformation may be regarded as a combination of linear transformations and an inversion.

5. Find the fixed points of the bilinear transformation (83).

Ans. $z = (a - d)/2c \pm \sqrt{[(a - d)/2c]^2 + (b/c)}.$

12. *Spherical representation of complex numbers*

Imagine a sphere of unit diameter tangent to the complex plane at the origin, Figure 1.20. Draw a line from the "north pole" N of the sphere to a point z in the plane. The second point z' of intersection of this line with the sphere may be taken to represent the point z in the plane. The unit circle in the plane is represented by the equator on the sphere; its radii are represented by the meridians; points inside the unit circle correspond to points in the southern hemisphere and points outside the unit circle are represented by points in the northern hemisphere. Thus two very unequal regions in the plane are represented by equal regions on the sphere.

This method of representation of a complex variable was used by Neumann in his treatise "Vorlesungen Über Riemanns Theorie der Abelschen Integrale" (Leipzig, Teubner, 2nd edition, 1884), and the

complex sphere is sometimes referred to as Neumann's sphere. The method is particularly helpful in thinking about functions of a complex variable when the variable is in distant parts of the complex plane. No matter in what direction we go to infinity, we approach the north pole on the Neumann sphere; point N is *the point at infinity*, $z = \infty$.

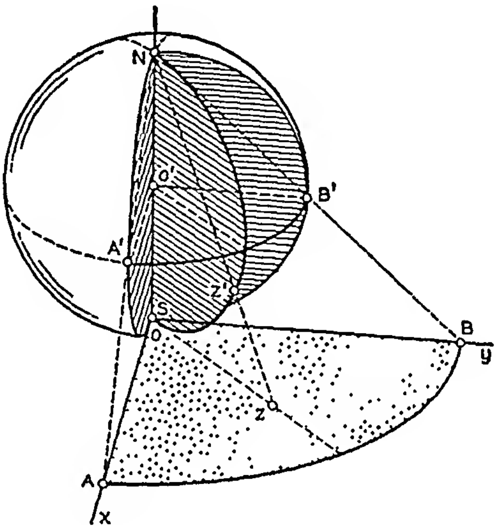

FIG. 1.20. The Neumann sphere for the representation of complex numbers.

On the other hand, there are obvious disadvantages. Without a spherical model, figures are not easy to draw; and one is forced to rely to a greater extent on imagination.

Problems

1. Show that

$$\sin\theta = \frac{2\rho}{1+\rho^2}, \qquad \cos\theta = \frac{\rho^2-1}{\rho^2+1},$$

where θ is the angle $NO'Z'$ in Figure 1.20 and ρ is the absolute value of z.

2. Show that circles on the Neumann sphere represent circles in the complex plane and vice versa.

References

1. C. Chrystal, *Algebra*, Parts I and II, A. & C. Black, Ltd., London.
2. E. W. Hobson, *A Treatise on Plane Trigonometry*, Cambridge University Press.

CHAPTER II

THEORY OF APPROXIMATION

1. *Linear interpolation*

Interpolation is a process of estimating the value of a function at an intermediate point when its values are known only at certain specified points. *Linear* interpolation is illustrated in Figure 2.1 where we assume that the curve $y = f(x)$ between two given points A, B can be approximated by the straight line joining them. The equation of the straight line is

$$y - y_1 = \frac{\Delta y}{\Delta x}(x - x_1), \tag{1}$$

where the *first differences* are

$$\Delta x = x_2 - x_1, \qquad \Delta y = y_2 - y_1. \tag{2}$$

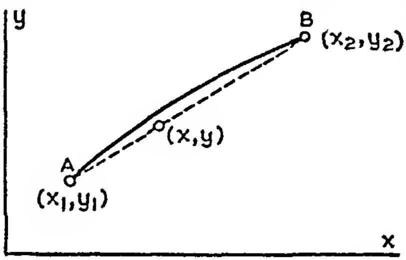

FIG. 2.1. In linear interpolation the true function represented by the curve AB is approximated by the straight line AB.

By direct substitution it is easy to verify that the values $x = x_1, y = y_1$ and $x = x_2, y = y_2$ actually satisfy (1).

The approximate value of the function y, at any point x in the interval (x_1,x_2), is taken as

$$y = y_1 + \frac{x - x_1}{\Delta x}\Delta y. \tag{3}$$

This is the most frequently used form of interpolation.

2. *Quadratic interpolation*

Our estimate of intermediate values would be better if we could allow somehow for the possible curvature of the function. With just two points we have no information about the curvature; the curve could be either convex or concave. With three points, however, we are better off; although we still have to decide on a particular method of taking the curvature into consideration. We could use, for instance, a suitable French curve and perform interpolation graphically. We could always draw a circle through three points. This would bring a certain uniformity into the interpolation; but the practical difficulty is that the circle is complicated to deal with analytically and difficult to draw accurately when the

curvature is small. And the curvature must be small or we should not try to interpolate. A practical way, from the computational point of view, is to obtain a parabola passing through the three given points, Figure 2.2. Just as the linear function

$$y = a + bx \tag{4}$$

is the basis of linear, or two-point interpolation, the quadratic function

$$y = a + bx + cx^2 \tag{5}$$

may be used for three-point interpolation. Suppose that we know the values

$$y = y_1, y_2, y_3$$

when

$$x = x_1, x_1 + h, x_1 + 2h.$$

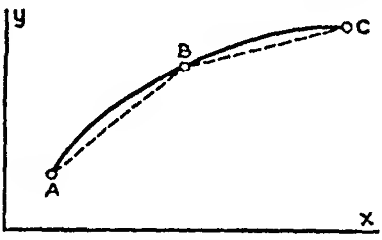

FIG. 2.2. In quadratic interpolation the true curve is approximated by a parabola (the solid line) passing through three points A, B, C on the true curve.

Extending the linear interpolation formula (3) by adding a quadratic term, we have

$$y = y_1 + \frac{x - x_1}{h}(y_2 - y_1) + k(x - x_1)(x - x_1 - h). \tag{6}$$

The last term is chosen deliberately so that it vanishes at $x = x_1$ and $x = x_1 + h$ and thus the curve still passes through the first two points. The constant k is chosen to make the curve pass through the third point $(x_1 + 2h, y_3)$

$$y_3 = y_1 + 2(y_2 - y_1) + 2kh^2, \qquad k = (y_3 - 2y_2 + y_1)/2h^2.$$

Thus we obtain the following quadratic interpolation formula

$$y = y_1 + \frac{x - x_1}{h}(y_2 - y_1) + \frac{(x - x_1)(x - x_1 - h)}{2h^2}(y_3 - 2y_2 + y_1). \tag{7}$$

The *second difference* $y_3 - 2y_2 + y_1$ is the difference $(y_3 - y_2) - (y_2 - y_1)$ of the adjacent first differences.

We could easily extend the formula by adding a cubic term

$$k(x - x_1)(x - x_1 - h)(x - x_1 - 2h) \tag{8}$$

and making the curve pass through the next point $(x_1 + 3h, y_4)$. We should find

$$k = (y_4 - 3y_3 + 3y_2 - y_1)/3!\,h^3. \tag{9}$$

The numerator is the third difference representing the difference of the adjacent second differences. The reader should not find it difficult now to obtain *Newton's interpolation formula* of the nth order.

Problems

1. On pages 666 and 668 of G. N. Watson's "A Treatise on the Theory of Bessel Functions" we find the following entries for the Bessel function of order zero and of the first kind, $J_0(x)$,

$$J_0(1) = 0.7651977, \qquad J_0(1.1) = 0.7196220,$$
$$J_0(1.2) = 0.6711327.$$

In this case $h = 0.1$ and (7) becomes

$$J_0(x) = J_0(1) - 0.455757(x-1) - 0.14568(x-1)(x-1.1).$$

Hence

$$J_0(1.04) = 0.7651977 - 0.0182303 + 0.0003496$$
$$= 0.7473170.$$

Watson's table contains $J_0(1.04) = 0.7473390$; the first four decimals are identical. The linear interpolation formula does not contain the third term and gives $J_0(1.04) = 0.7469674$.

2. Make calculations similar to the foregoing and check the relative accuracy of interpolation formulas of various orders by comparison with the entries in the table.

3. Derive Newton's Interpolation Formula of the nth Order.

Ans.
$$f(x) = f(a) + \frac{x-a}{h}\Delta f(a) + \frac{(x-a)(x-a-h)}{2!\,h^2}\Delta^2 f(a)$$
$$+ \frac{(x-a)(x-a-h)(x-a-2h)}{3!\,h^3}\Delta^3 f(a) + \cdots$$
$$+ \frac{(x-a)(x-a-h)\cdots(x-a-\overline{n-1}h)}{n!\,h^n}\Delta^n f(a),$$

where the successive differences are

$$\begin{aligned}
\Delta f(a) &= f(a+h) - f(a),\\
\Delta^2 f(a) &= \Delta[\Delta f(a)] = \Delta f(a+h) - \Delta f(a),\\
\Delta^3 f(a) &= \Delta[\Delta^2 f(a)] = \Delta^2 f(a+h) - \Delta^2 f(a),\\
&\cdots\cdots\cdots\\
\Delta^n f(a) &= \Delta[\Delta^{n-1} f(a)] = \Delta^{n-1} f(a+h) - \Delta^{n-1} f(a).
\end{aligned}$$

3. *Lagrange's formula*

A much more obvious form of a polynomial of the $(n-1)$th degree which assumes the values $y_1, y_2, \cdots y_n$ when x takes on the values x_1,

$x_2, \cdots x_n$ is Lagrange's form

$$
\begin{aligned}
y = {} & \frac{(x - x_2)(x - x_3)\cdots(x - x_n)}{(x_1 - x_2)(x_1 - x_3)\cdots(x_1 - x_n)}\, y_1 \\
& + \frac{(x - x_1)(x - x_3)\cdots(x - x_n)}{(x_2 - x_1)(x_2 - x_3)\cdots(x_2 - x_n)}\, y_2 + \cdots \qquad (10) \\
& + \frac{(x - x_1)(x - x_2)\cdots(x - x_{n-1})}{(x_n - x_1)(x_n - x_2)\cdots(x_n - x_{n-1})}\, y_n.
\end{aligned}
$$

Each term is of the $(n - 1)$th degree. When $x = x_1$ the first term is equal to y_1 and the remaining terms vanish; etc.

Lagrange's formula is not convenient for the purposes of interpolation because all terms are of the same order of magnitude instead of becoming progressively smaller as in Newton's formula; but when one wishes to approximate a given function by a polynomial, the above equation gives the most direct answer. Of course, after expansion into powers of x, Newton's and Lagrange's formulas become identical.

Problem. Approximate $y = (1 + x)^{-1}$ in the interval (0,2) by a sequence of Lagrange's polynomials passing through equispaced points; that is, passing through $x = 0, 2$; then through $x = 0, 1, 2$; then through $x = 0, 2/3, 4/3, 2$; etc.

Ans. $y_1(x) = 1 - \frac{1}{3}x, \qquad y_2(x) = 1 - \frac{2}{3}x + \frac{1}{6}x^2,$

$y_3(x) = 1 - \frac{89}{105}x + \frac{3}{7}x^2 - \frac{3}{35}x^3.$

4. *Approximation by power series*

Interpolation is based on approximation of functions by polynomials whose values coincide with those of the given function at a *specified number of points*. Another type of approximation is *approximation in the vicinity of a given point*.

Suppose we wish to approximate the curve in Figure 2.1 by a straight line in the vicinity of the point A. We shall certainly want the straight line to pass through A. Next we observe that if B is allowed to move nearer to A, the line AB will be a better approximation in the vicinity of A. *The tangent at A* conforms to our idea of the best straight line approximation in the vicinity of A; thus the values of the given and approximating functions and the values of their first derivatives should be equal at A.

We shall now generalize. Let us take the following polynomial of the nth degree

$$f(x) = A_0 + A_1(x - a) + A_2(x - a)^2 + \cdots + A_n(x - a)^n \qquad (11)$$

and obtain its successive derivatives

$$\begin{aligned} f'(x) &= A_1 + 2A_2(x-a) + 3A_3(x-a)^2 + \cdots + nA_n(x-a)^{n-1}, \\ f''(x) &= 1\cdot 2A_2 + 2\cdot 3A_3(x-a) + \cdots + (n-1)nA_n(x-a)^{n-2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ f^{(n)}(x) &= 1\cdot 2\cdot 3\cdots nA_n, \quad f^{(n+1)}(x) = 0. \end{aligned} \tag{12}$$

The values of the function and its derivatives at $x = a$ are

$$\begin{aligned} &f(a) = A_0, \quad f'(a) = A_1, \quad f''(a) = 1\cdot 2A_2, \\ &f'''(a) = 1\cdot 2\cdot 3A_3, \cdots f^{(n)}(a) = n!\,A_n, \quad f^{(n+1)}(a) = 0. \end{aligned} \tag{13}$$

Thus the coefficients of the polynomial can be expressed in terms of its value and the values of its derivatives at a given point $x = a$

$$\begin{aligned} f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) \\ + \frac{(x-a)^3}{3!}f'''(a) + \cdots + \frac{(x-a)^n}{n!}f^{(n)}(a). \end{aligned} \tag{14}$$

If $f(x)$ is a given function, although not a polynomial, the above equation yields an approximation in the vicinity of $x = a$. Formally, we can take n as large as we please and obtain *Taylor's infinite series* for $f(x)$; but then we shall have to investigate the convergence of the series.

It should never be forgotten that so far we have been concerned only with functions of a real variable and with best polynomial approximations in the vicinity of a fixed point; without further investigation no conclusion can be drawn as to how good this approximation will be at some distance from the given point. For instance, the function

$$f(x) = \exp(-1/x^2) \tag{15}$$

and all its derivatives vanish at $x = 0$. The approximating function is $F(x) = 0$. When x is very small, there is no question that zero is a good approximation to (15); but certainly the approximation is bad if $x = 1$. We cannot improve the polynomial because we get $F(x) = 0$ no matter how many terms we include.

At first this phenomenon may seem strange; but the mystery will disappear after studying the behavior of power series for *complex values* of the variable (Chapter 4). Here we shall only note that if $x = iu, f(x) = \exp(1/u^2)$; and as u approaches zero, $f(x)$ becomes infinite. On the other hand, power series do not behave in this way.

Problem. Approximate $y = (1+x)^{-1}$ in the vicinity of $x = 0$ by a sequence of Taylor's polynomials

$$\begin{aligned} &y_0(x) = 1, \qquad y_1(x) = 1 - x, \qquad y_2(x) = 1 - x + x^2, \\ &y_3(x) = 1 - x + x^2 - x^3, \text{ etc.} \end{aligned}$$

5. *Approximation "on the average"*

In the preceding sections we have obtained two approximations of a given function by polynomials of the nth degree: Newton's and Taylor's. Of the two, the first is obviously the better throughout a given interval; but it is not necessarily the best except at points where the function and its approximation coincide.

For example, if we wished to approximate a given function $f(x)$ in a given interval $a \leq x \leq b$ by a constant A, that is, by a straight line parallel to the x-axis, Figure 2.3, we would not draw the approximating line MN through either of the end points A, B. More probably we would approximate $f(x)$ by its average value in the interval, so that the integral of the positive error would just cancel the integral of the negative error.

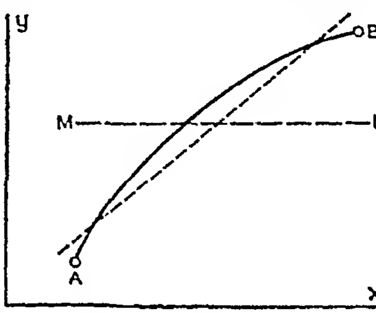

Fig. 2.3. Illustration of approximations "on the average."

Suppose we wish to improve the approximation by making the approximating line inclined. We now have two parameters at our disposal: the height of the middle point and the slope of the straight line or, in the algebraic form, the coefficients of the linear function $A_0 + A_1x$. How are we to choose these parameters? One way is to minimize the integrated square of the error

$$I = \int_a^b [f(x) - A_0 - A_1x]^2\, dx. \tag{16}$$

It is a uniform method of determination of the coefficients and one which applies equally well to an approximating polynomial of any degree, in which case we would minimize

$$I = \int_a^b [f(x) - A_0 - A_1x - \cdots - A_nx^n]^2\, dx. \tag{17}$$

It is clear that the method is arbitrary. Why should we minimize the square of the error rather than some other power or even some more complicated function of the error? The odd powers cannot be used for then I has no minimum; but any even power could serve the purpose except that the evaluation of the coefficients would become very complicated. Of course, the fourth power law would tend to suppress really big errors at the expense of smaller errors to a greater extent than the square law. The latter treats all errors more equally.

Another point: We may wish the approximating function to be better in some particular subinterval; we may decide, for example, that we would be satisfied with a poorer approximation near the ends of the interval or near the middle. This objective can be easily achieved if we include a *weight function* $W(x)$ and choose the coefficients by minimizing the integral of the weighted square of the error

$$I = \int_a^b W(x)[f(x) - A_0 - A_1 x - \cdots - A_n x^n]^2 \, dx, \tag{18}$$

where $W(x) \geq 0$ but not identically equal to zero.

Let us see how the method actually works when $W(x) = 1$. By squaring the integrand and evaluating whatever integrals we can, we obtain from (16)

$$\begin{aligned} I = \int_a^b [f(x)]^2 \, dx + A_0^2(b - a) + \tfrac{1}{3}A_1^2(b^3 - a^3) \\ - 2A_0 \int_a^b f(x) \, dx - 2A_1 \int_a^b xf(x) \, dx + A_0 A_1(b^2 - a^2). \end{aligned} \tag{19}$$

The requirements for the minimum of this function are

$$\frac{\partial I}{\partial A_0} = 0, \qquad \frac{\partial I}{\partial A_1} = 0. \tag{20}$$

Differentiating (19), first with respect to A_0 and then with respect to A_1, we obtain

$$\begin{aligned} 2A_0(b - a) + A_1(b^2 - a^2) - 2\int_a^b f(x) \, dx = 0, \\ \tfrac{2}{3}A_1(b^3 - a^3) + A_0(b^2 - a^2) - 2\int_a^b xf(x) \, dx = 0. \end{aligned} \tag{21}$$

These equations are linear in the unknown coefficients; they would still be linear in the more general cases (17) and (18).

Now that the general scheme for selecting the coefficients of the approximating polynomial is clear, let us organize the procedure in order to simplify numerical work. It would certainly be advantageous to deal with a standard interval. By choosing the mid-point of the interval (a,b) as the new origin and changing the scale, the interval can always be made $(-1,1)$; hence we introduce a new independent variable ξ defined by

$$\xi = \frac{2}{b - a}\left(x - \frac{a + b}{2}\right), \qquad x = \tfrac{1}{2}(b - a)\xi + \tfrac{1}{2}(b + a). \tag{22}$$

Thus, without loss of generality we may let $a = -1$, $b = 1$ in (16). Next,

instead of using the powers of x as the elements of the approximating function, we introduce a set of polynomials

$$p_0(x), \quad p_1(x), \quad p_2(x), \cdots, p_n(x) \tag{23}$$

so chosen that the calculation of the coefficients becomes easier. At this stage we shall subject these polynomials to only one condition:

$$p_m(x) \textit{ is a polynomial of } m\text{th } \textit{degree}. \tag{A}$$

Hence, we can rewrite (17) as follows

$$I = \int_{-1}^{1} [f(x) - a_0p_0(x) - a_1p_1(x) - \cdots - a_np_n(x)]^2\,dx. \tag{24}$$

We have not changed the degree of the approximating polynomial but merely reshuffled the various powers into special groups. Squaring the integrand, we have

$$\begin{aligned} I = \int_{-1}^{1} [f(x)]^2\,dx &+ \sum_{m=0}^{n} a_m^2 \int_{-1}^{1} [p_m(x)]^2\,dx \\ &- 2\sum_{m=0}^{n} a_m \int_{-1}^{1} f(x)p_m(x)\,dx + 2\sum_m\sum_k{}' a_ma_k \int_{-1}^{1} p_m(x)p_k(x)\,dx. \end{aligned} \tag{25}$$

The prime that goes with the double summation is to remind us that $m \neq k$, since the terms corresponding to $m = k$ have been exhibited separately. Suppose now that we have been successful in finding p-functions such that

$$\int_{-1}^{1} p_m(x)p_k(x)\,dx = 0, \quad \text{if } m \neq k; \tag{B}$$

$$= 1, \quad \text{if } m = k. \tag{C}$$

Equation (25) would then become simplified

$$I = \sum_{m=1}^{n} a_m^2 - 2\sum_{m=0}^{n} a_m \int_{-1}^{1} f(x)p_m(x)\,dx + \int_{-1}^{1} [f(x)]^2\,dx. \tag{26}$$

The partial derivative with respect to a_m must vanish in order to make I minimum; thus,

$$\frac{\partial I}{\partial a_m} = 2a_m - 2\int_{-1}^{1} f(x)p_m(x)\,dx = 0. \tag{27}$$

Hence,

$$a_m = \int_{-1}^{1} f(x)p_m(x)\,dx \tag{28}$$

is obtained by integration without the necessity of solving $(n + 1)$ linear equations in $(n + 1)$ unknowns, as we should have to do if we dealt directly with the original polynomial in (17).

All that now remains is to calculate the set of polynomials satisfying the conditions A, B, C. This can be done step by step. Since $p_0(x)$ is of zero degree, we write $p_0(x) = a$ and determine this constant from

$$\int_{-1}^{1} a^2\,dx = 1, \qquad a = \sqrt{1/2}. \tag{29}$$

The algebraic sign of a is not uniquely determined by the conditions A, B, C; and we are at liberty to choose it to suit our convenience. This will also be true of the sign associated with $p_n(x)$. Next we determine $p_1(x) = a + bx$ from

$$\int_{-1}^{1} p_0(x)p_1(x)\,dx = \frac{1}{\sqrt{2}}\int_{-1}^{1} (a + bx)\,dx = 0,$$

$$\int_{-1}^{1} [p_1(x)]^2\,dx = \int_{-1}^{1} (a + bx)^2\,dx = 1.$$

Solving for a and b, we find

$$p_1(x) = \sqrt{3/2}\,x. \tag{30}$$

Then we proceed to the next polynomial, and next, and next. The results are usually presented in the following form

$$p_m(x) = \sqrt{(2m + 1)/2}\,P_m(x), \tag{31}$$

where

$$\begin{aligned}
&P_0(x) = 1, \qquad P_1(x) = x, \qquad P_2(x) = \tfrac{1}{2}(3x^2 - 1),\\
&P_3(x) = \tfrac{1}{2}(5x^3 - 3x), \qquad P_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3),\\
&P_5(x) = \tfrac{1}{8}(63x^5 - 70x^3 + 15x),\\
&P_6(x) = \tfrac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5),\\
&P_7(x) = \tfrac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x).
\end{aligned} \tag{32}$$

The P-polynomials are called *Legendre polynomials*. Functions satisfying condition (B) are called *orthogonal;* those satisfying (C) are *normalized.* Legendre polynomials are orthogonal but not normalized. The constant multipliers in (31) are *normalization factors*.

Problem. Approximate $y = (1 + x)^{-1}$ by a sequence of Legendre's approximations in the interval $0 \le x \le 2$.

Ans.
$$\begin{aligned}
y_1(x) &= \tfrac{1}{2}\log 3 + 3(1 - \log 3)P_1(x - 1) = 0.845 - 0.296x\\
y_2(x) &= y_1(x) + (-15 + \tfrac{55}{4}\log 3)P_2(x - 1)\\
&= 0.951 - 0.614x + 0.159x^2\\
y_3(x) &= y_2(x) + (\tfrac{196}{3} - \tfrac{119}{2}\log 3)P_3(x - 1)\\
&= 0.985 - 0.818x + 0.415x^2 - 0.085x^3.
\end{aligned}$$

6. *Fourier approximation*

In the preceding sections we have considered various polynomial approximations of a given function. One set of polynomials was designed to approximate the function in the vicinity of a given point, another set to represent the function exactly for specified values of the independent variable, and still another set to approximate the function in a given interval without any strings attached as to the disposition of points of exact coincidence. In the next section we shall compare these three types of approximation. What we wish to consider here is the use of functions other than polynomials. Undoubtedly the student has already begun to suspect that sets of functions other than polynomials might be used for the purpose. This is actually the case; there are infinitely many ways of representing a given function, some much more suitable than others for a particular purpose. This latter point will be illustrated when we come to the solution of ordinary and partial differential equations. Here we shall confine ourselves to the Fourier representation of a function in terms of sines and cosines of integral multiples of the independent variable.

The representation is of the following type

$$\begin{aligned} f(x) \simeq \tfrac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx \\ + b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx. \end{aligned} \tag{33}$$

The right side is periodic with the period 2π; hence we cannot hope to represent $f(x)$ in an interval of greater length than 2π unless $f(x)$ is also periodic with the same period. We may be able to represent $f(x)$ in the interval $(-\pi,\pi)$ or $(0,2\pi)$ or $(x_0,x_0 + 2\pi)$, where x_0 is some particular value of x; but not in the interval $(0,4\pi)$ — at least not by an expression (33).

How are we going to determine the coefficients? One way would be to make equation (33) exact for $(2n + 1)$ preassigned values of x in the chosen interval $(x_0,x_0 + 2\pi)$. This would necessitate solution of $(2n + 1)$ linear equations with $(2n + 1)$ unknowns — a difficult numerical task when n is large (except when the subintervals are equal). There is a much simpler way which is based on the following orthogonality property

$$\int_{x_0}^{x_0+2\pi} \cos mx \cos kx \, dx = 0, \quad \text{if } m \neq k,$$

$$\int_{x_0}^{x_0+2\pi} \sin mx \sin kx \, dx = 0, \quad \text{if } m \neq k,$$

$$\int_{x_0}^{x_0+2\pi} \cos mx \sin kx \, dx = 0.$$

Thus if we multiply both sides of (33) by cos mx or sin mx and integrate from $x = x_0$ to $x = x_0 + 2\pi$ we obtain

$$\begin{aligned} \int_{x_0}^{x_0+2\pi} f(x) \cos mx\, dx &= a_m \int_{x_0}^{x_0+2\pi} \cos^2 mx\, dx, \\ \int_{x_0}^{x_0+2\pi} f(x) \sin mx\, dx &= b_m \int_{x_0}^{x_0+2\pi} \sin^2 mx\, dx. \end{aligned} \tag{35}$$

Since

$$\int_{x_0}^{x_0+2\pi} \cos^2 mx\, dx = \int_{x_0}^{x_0+2\pi} \sin^2 mx\, dx = \pi, \quad \text{if } m \neq 0,$$

we have

$$\begin{aligned} a_m &= \frac{1}{\pi} \int_{x_0}^{x_0+2\pi} f(x) \cos mx\, dx, \\ b_m &= \frac{1}{\pi} \int_{x_0}^{x_0+2\pi} f(x) \sin mx\, dx. \end{aligned} \tag{36}$$

Because of the factor 1/2 associated with a_0 in (33) the first of these equations is true also when $m = 0$.

The coefficients have thus been determined independently of each other. There can be no question that if (33) were exact, the coefficients would be given by (36); but if there is no prior evidence that $f(x)$ can be represented by an expression of the form (33), we cannot be sure that our scheme yields the correct representation of $f(x)$. This should be obvious since there is nothing in our method of obtaining the coefficients that would tell us how many terms should be included and the method would work just as well if we inadvertently omitted cos $2x$ or all the sine functions. Suppose we let n be infinite but failed to notice that the right side of (33) has the period 2π and performed our integration from $x = x_0$ to $x = x_0 + 4\pi$. We would still obtain the coefficients of a series of circular functions; but this series would not represent $f(x)$ even approximately except in very special cases. Hence, it is essential to prove that the infinite series whose coefficients are given by (36) really does represent $f(x)$. The proof is beyond the scope of this book and we shall merely state that if $f(x)$ is single-valued and piece-wise* continuous in the interval $(0,2\pi)$, then as n increases indefinitely, the function can be approximated by (33) with an increasing degree of accuracy everywhere in the interval except at points of discontinuity. At the latter points the series converges to the mean value of the limits approached by the series when points of discontinuity are approached from left and right.

* A function is piece-wise continuous in an interval (a,b) if (a,b) can be subdivided into a finite number of subintervals in which the function is continuous.

Problems

1. Expand the following functions in Fourier series:

a) $f(x) = -1, \quad -\pi < x < 0;$
$\quad\;\; = 1, \quad 0 < x < \pi.$

b) $f(x) = x, \quad -\pi < x < \pi.$

c) $f(x) = x, \quad 0 < x < 2\pi.$

Ans. a) $f(x) = \dfrac{4}{\pi}\sum_{n=0}^{\infty} \dfrac{1}{2n+1} \sin (2n+1)x,$

b) $f(x) = 2\sum_{n=1}^{\infty} \dfrac{(-)^{n-1}}{n} \sin nx,$

c) $f(x) = \pi - 2\sum_{n=1}^{\infty} \dfrac{\sin nx}{n}.$

2. Show that if the interval is (a,b), the Fourier expansion of $f(x)$ is

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{2\pi nx}{b-a} + \sum_{n=1}^{\infty} b_n \sin \frac{2\pi nx}{b-a},$$

$$a_n = \frac{2}{b-a}\int_a^b f(x) \cos \frac{2\pi nx}{b-a}\, dx,$$

$$b_n = \frac{2}{b-a}\int_a^b f(x) \sin \frac{2\pi nx}{b-a}\, dx.$$

3. Obtain the Fourier expansions given on page 138 of "Smithsonian Mathematical Formulae and Tables of Elliptic Functions."

7. *In retrospect*

A comparison between various polynomial approximations to a given function is very instructive. Figures 2.4, 2.5, and 2.6 show the function $y = (1 + x)^{-1}$ and some approximations to it in the interval $0 \leq x \leq 2$. On the scale used, the cubic approximations of the Lagrange and Legendre types are too close to the exact curve to be conveniently exhibited.

The power series approximations are seen to be in complete distress except for small values of x. The answer to the mystery lies, as explained in Chapter 4, in the sensitivity of power series to singularities that may exist outside the given interval. In the present example, the singularity is at $x = -1$; but the power series for $y = (1 + x^2)^{-1}$ would behave very similarly on account of the singularities at $x = \pm i$.

The resemblance between power series and a series of Legendre polynomials is very superficial as the graphical illustrations show.

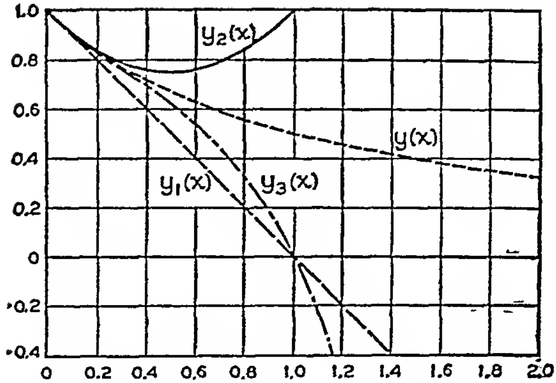

FIG. 2.4. The function $y(x) = (1 + x)^{-1}$ and its successive approximations by power series in the vicinity of $x = 0$ (see the problem of Section 4). Note how poor the approximations are for the larger values of x.

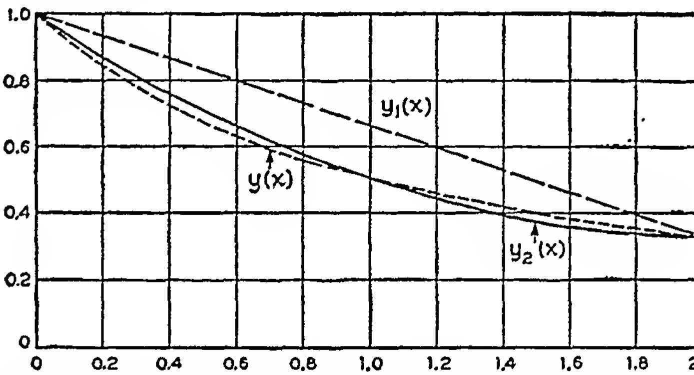

FIG. 2.5. The function $y(x) = (1 + x)^{-1}$ and its successive approximations by polynomials when the approximations are made perfect at specified points in the interval $0 \leq x \leq 2$ (see the problem of Section 3).

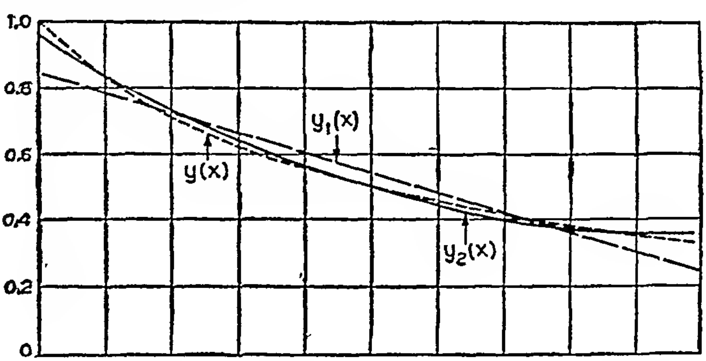

FIG. 2.6. The function $y(x) = (1 + x)^{-1}$ and its successive approximations "on the average" by Legendre polynomials (see the problem of Section 5). The crossings of the exact and approximate curves are not specified; instead, the integral of the square of the error is minimized.

analytic difference is: each time a new term is added to the power series, the coefficients of the previous powers remain unaltered; but each time a new Legendre polynomial is added, the coefficients of *all* previous powers are altered.

Approximations of our function by powers of x do not extend very far from $x = 0$; but for small values of x they are equally good if x is negative or complex. The series of Legendre approximations, however, will show signs of distress if it is used outside the range of x for which it has been derived.

Problems

1. Find the quadratic approximation to $y = \cos(\pi x/2)$ with the least mean square deviation in the interval $-1 \leq x \leq 1$.

Ans. $\dfrac{3}{\pi}\left(\dfrac{20}{\pi^2} - 1\right) + \dfrac{15}{\pi}\left(1 - \dfrac{12}{\pi^2}\right)x^2.$

2. Find the linear approximations to the following functions

a) $y = \sin x,\quad 0 \leq x \leq \pi/2;$ b) $y = \sin x,\quad \pi/2 \leq x \leq \pi;$
c) $y = \cos x,\quad 0 \leq x \leq \pi/2;$ d) $y = \cos x,\quad \pi/2 \leq x \leq \pi;$

with the condition that the mean square error in the designated intervals be minimum.

Ans. If $A = \dfrac{8}{\pi}\left(1 - \dfrac{3}{\pi}\right),\quad B = \dfrac{24}{\pi^2}\left(\dfrac{4}{\pi} - 1\right)$; then

a) $A + Bx$, b) $A - B(x - \pi)$, c) $A - B\left(x - \dfrac{\pi}{2}\right)$, d) $-A - B\left(x - \dfrac{\pi}{2}\right)$.

3. For exercises in the expansion of functions in power series and Fourier series see "Smithsonian Mathematical Formulae and Tables of Elliptic Functions," Chapter VI.

CHAPTER III

Solution of Equations

1. *A general method of solution of algebraic and transcendental equations*

The most straightforward and general method of solving algebraic and transcendental equations is the *graphical method* or its arithmetic counterpart, the *table method.* To find the real roots of

$$F(x) = f(x), \tag{1}$$

we merely plot two curves

$$y = F(x), \qquad y = f(x), \tag{2}$$

and find the points of intersection. To increase the accuracy the curves are replotted on a larger scale in the immediate vicinity of each point of intersection. Actual plotting of the curves is not essential; tables of $F(x)$ and $f(x)$ will suffice.

To obtain complex roots we let $x = u + iv$ and write

$$\begin{aligned} F(x) &= A(u,v) + iB(u,v), \\ f(x) &= C(u,v) + iD(u,v); \end{aligned} \tag{3}$$

then we substitute in (1) and equate the real and imaginary parts

$$A(u,v) = C(u,v), \qquad B(u,v) = D(u,v). \tag{4}$$

Finally we plot v vs. u from each equation and find the points of intersection.

2. *Examples — real roots*

As the first example let us take the following equation

$$\cos x - \frac{\sin x}{x} = 0. \tag{5}$$

This equation occurs in the theory of natural oscillations of a circular membrane clamped along the circumference and a radius as shown in Figure 3.1; in the theory of natural oscillations of air inside a rigid spherical envelope; and in the theory of electromagnetic waves in a wave guide whose cross section is that shown in Figure 3.1. If a is the radius, the wavelengths of oscillations are found from $\lambda = 2\pi a/x$, where x is a root of (5).

Instead of plotting $\cos x$ and $(\sin x)/x$ directly, let us transform (5) into

$$\tan x = x; \tag{6}$$

then we need only a curve for $y = \tan x$ and the straight line $y = x$. Since changing the sign of x leaves (6) unaltered, to every positive root there corresponds a negative root of the same magnitude. Figure 3.2 shows that the positive roots are somewhat smaller than $(n + \frac{1}{2})\pi$, $n > 0$. Starting with the smallest root, we plot* the region in the vicinity of the root near $x = 3\pi/2$ on a larger scale, Figure 3.3. Thus for the smallest positive root we obtain $x = 4.4934$.

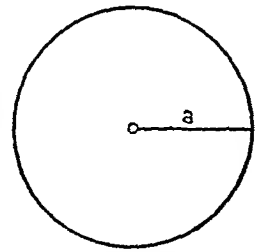

Fig. 3.1. A circular membrane clamped along the circumference and a radius.

Another equation

$$\sin x + \frac{\cos x}{x} - \frac{\sin x}{x^2} = 0 \tag{7}$$

comes from the theory of natural electrical oscillations in a spherical cavity resonator. If a is the radius of the cavity, the natural wavelength and frequency are $\lambda = 2\pi a/x$ and $f = 3 \times 10^{10}/\lambda$ (if λ is in centimeters). Again we transform the above equation into a simpler form

$$\tan x = \frac{x}{1 - x^2}. \tag{8}$$

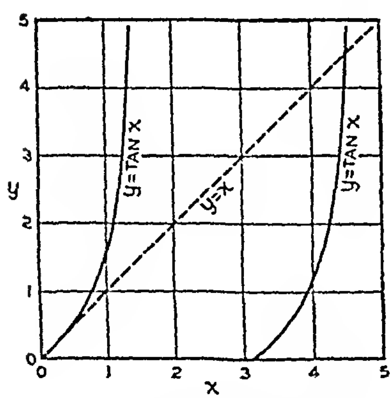

Fig. 3.2. Illustration of the graphical method of solving equation (6).

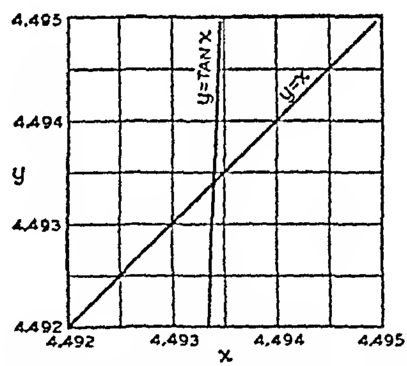

Fig. 3.3. A large scale plot of the region around the second intersection of the curves in Figure 3.2.

Figure 3.4 shows the exploratory plot of the two sides of this equation which yields rough approximations to the smallest roots; replotting the region around the smallest positive root we obtain $x = 2.74371$.

* Hayashi table of circular, exponential and hyperbolic functions.

As a last illustration let us take an equation involving Bessel functions

$$J_0(\beta a)N_0(\beta b) - J_0(\beta b)N_0(\beta a) = 0. \tag{9}$$

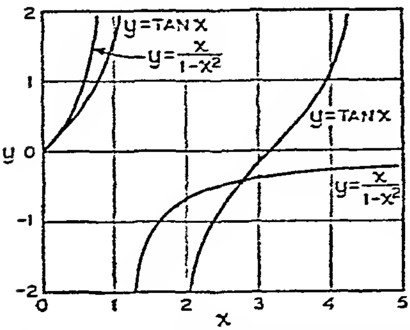

FIG. 3.4. An exploratory plot of the two sides of equation (8) to determine approximate locations of the roots.

This equation occurs in the theory of free vibrations of an annular membrane and in the theory of propagation of sound and electromagnetic waves between two coaxial cylinders. The radii of the cylinders are a and b; $\beta = 2\pi/\lambda$ where λ is the wavelength; J_0 and N_0 are Bessel functions of order zero.*

In the present case we have one set of roots for each ratio b/a. If we are interested in exploring a large range of this parameter, we shall profit by writing $\beta a = p$, $\beta b = q$ and transforming (9) into

$$\frac{J_0(p)}{N_0(p)} = \frac{J_0(q)}{N_0(q)}. \tag{10}$$

Figure 3.5 represents either side of this equation. If the wavelength λ, and hence the "phase constant" β, as well as the radius a of the inside cylinder are known, p is known. This p determines the ordinate and, therefore, the permissible values of q. Hence we obtain the permissible values of the radius of the outside cylinder.

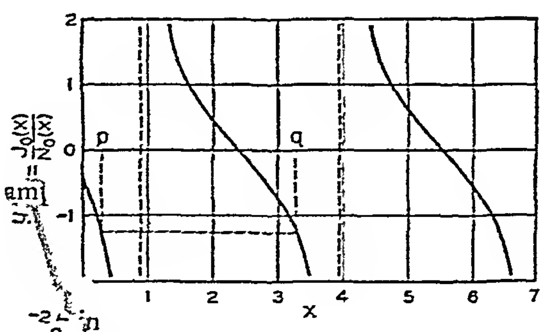

FIG. 3.5. The m... aster curve for obtaining the roots of equation (10).

To cover a wide range of parameters we may start with the first two branches of the curve in Figure 3.5, pick out pairs (p,q) corresponding to equal ordinates, and plot q vs. p, Figure 3.6. The curve can be continued

* In the Jahnke and Emde notation; G. N. Watson uses Y_0 instead of N_0.

by sliding along the next two branches of the master curve, etc. Finally we add a family of straight lines $q = (b/a)p$ so that $2\pi a/\lambda$ and $2\pi b/\lambda$ corresponding to various ratios of the radii can be read directly.

The q–p curve corresponding to the second "mode" of oscillation or propagation is obtained by selecting pairs (p,q) from the first and third, second and fourth, etc., branches of the master curve in Figure 3.5.

It is obvious that the graphical or table method can be applied to almost any practical problem. Its usefulness might be limited if no tables of the functions involved in the equation to be solved were available, when the main problem might prove to be the preparation of such tables.

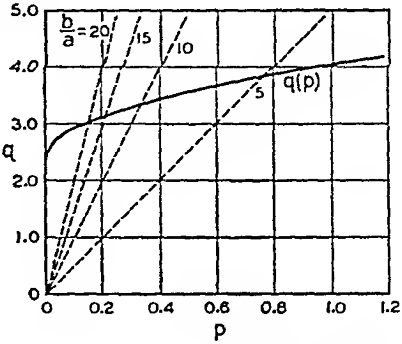

FIG. 3.6. The plot of q vs. p as obtained from Figure 3.5, and auxiliary straight lines for solving equation (9).

Problems

1. Find graphically the real root of $x^3 - 1.5x + 1 = 0$. Do it first by plotting $y = x^3 - 1.5x + 1$ and determining its intersection with the x-axis; then plot $y = x^3$ and $y = 1.5x - 1$. Note that the second method makes it easy to deal with a more general cubic equation $x^3 - px - q = 0$. Only one cubic curve $y = x^3$ is needed; the straight lines $y = px + q$ can be constructed from a pair of points. Show that if the cubic equation contains the square of the unknown, this term can be eliminated by a simple substitution $x = t + a$, where a is chosen to make the coefficient of t^2 equal to zero.

2. Calculate the two smallest positive roots of $\tan x = 2x^2$.

3. Let $y = f(x)$ be a continuous and single-valued function in the vicinity of a particular root of $f(x) = 0$. Suppose that $y_1 = f(x_1)$ and $y_2 = f(x_2)$ are of opposite signs.

Show that the root can be steadily approached by taking $x = x_1 - y_1 \dfrac{x_2 - x_1}{y_2 - y_1}$ and pairing it either with x_1 or x_2, depending on whether y has a sign opposite to y_1 or to y_2.

This is the interpolation method of solution.

3. *Examples — complex roots*

Consider the following equation

$$z^3 + z + 1 = 0. \tag{11}$$

Substituting $z = x + iy$ and equating to zero the real and imaginary

parts, we obtain

$$x^3 - 3xy^2 + x + 1 = 0, \tag{12}$$

$$y(3x^2 - y^2 + 1) = 0. \tag{13}$$

An obvious solution of (13) is $y = 0$. This would make z real and presumably we have already obtained the real roots of (11) as in the preceding section. For complex roots $y \neq 0$ and can be canceled from (13)

$$3x^2 - y^2 + 1 = 0. \tag{14}$$

The general procedure from now on would be to plot (12) and (14) and then locate the points of intersection. In the present case, however, it is simpler to solve (14) for y,

$$y^2 = 3x^2 + 1, \qquad y = \pm\sqrt{3x^2 + 1} \tag{15}$$

and substitute in (12),

$$-8x^3 - 2x + 1 = 0, \qquad x^3 = -\tfrac{1}{4}x + \tfrac{1}{8}.$$

In this equation x is real by definition and we proceed as in the preceding section. Thus we obtain $x = 0.34116$. Substituting in (15) we have $y = \pm 1.16154$ and the required complex roots of (11) are $z = 0.34116 \pm i1.16154$.

As our next example we take an equation from electromagnetic theory*

$$w \tanh w = -ik, \tag{16}$$

in which $w = u + iv$ is the unknown and k is a known positive parameter. To solve this equation we must first be able to separate the real and imaginary parts of the complex function $\tanh(u + iv)$. Referring to Chapter 9 we find

$$\tanh w = \frac{\sinh w}{\cosh w},$$

$$\begin{aligned} \sinh(u + iv) &= \sinh u \cos v + i \cosh u \sin v, \\ \cosh(u + iv) &= \cosh u \cos v + i \sinh u \sin v. \end{aligned} \tag{17}$$

Multiplying (16) by $\cosh w$, substituting from (17) and separating the real and imaginary parts, we have

$$\begin{aligned} u \sinh u \cos v - v \cosh u \sin v &= k \sinh u \sin v, \\ v \sinh u \cos v + u \cosh u \sin v &= -k \cosh u \cos v. \end{aligned} \tag{18}$$

Eliminating k, we obtain

$$u \sinh 2u = v \sin 2v. \tag{19}$$

* S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand Co., Inc., New York, 1943, p. 485.

Since u is real by definition, the left side of this equation is positive* and v must lie within the following intervals (such that $\sin 2v$ is positive or negative depending on whether v is positive or negative)

$$\begin{aligned} &0 \le v \le \pi/2, \qquad \pi \le v \le 3\pi/2, \cdots \; n\pi \le v \le (2n+1)\pi/2, \\ &0 \ge v \ge -\pi/2, \quad -\pi \ge v \ge -3\pi/2, \cdots -n\pi \ge v \ge -(2n+1)\pi/2. \end{aligned} \tag{20}$$

A reversal of the sign of w does not alter (16) and to every solution $u + iv$ there corresponds another solution $-u - iv$. Hence without loss of generality we can assume that v is positive. Since the algebraic sign of u for positive v is not determined by (19), we obtain it from the fact that k is known to be positive. From the second line† of (18) we have

$$k = -u \tan v - v \tanh u. \tag{21}$$

When v is defined by the first line of (20), $\tan v$ is positive and k can be positive only if u is negative. We now plot u vs. v from (19) and obtain a relationship between the real and imaginary parts of w which is entirely independent‡ of k. For the interval $0 \le v \le \pi/2$ the curve is that shown

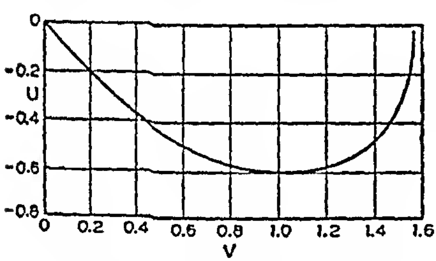

FIG. 3.7. The plot of the real part of a root of equation (16) vs. the imaginary part.

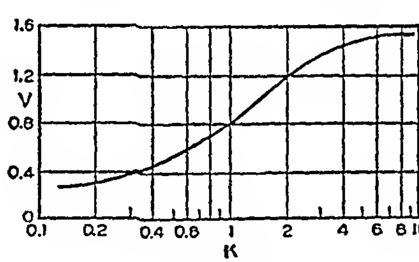

FIG. 3.8. The plot of the imaginary part of w vs. k in equation (16).

in Figure 3.7. Using this curve in conjunction with (21), we plot k vs. v, Figure 3.8. Thus for each value of k we can find v; and knowing v, we can find u from Figure 3.7.

Numerical work involved in computing complex roots is at best time consuming. There are excellent tables of hyperbolic and circular functions; but even so we have to find a real root of an equation just to obtain a point on the curve of Figure 3.7. The simplest procedure is to plot $y = u \sinh 2u$ and $y = v \sin 2v$ so that u and v are both abscissas; pairs (u,v) are then picked to correspond to equal ordinates.

* The function $\sinh 2u$ is odd so that $\sinh(-2u) = -\sinh 2u$.

† The first line does not furnish the required information as readily as the second.

‡ This is really a consequence of k having a constant phase (zero in the present case).

Problems

1. Compute the complex roots of

$$\text{a)}\quad z^3 + 3z - 2 = 0, \qquad \text{b)}\quad z^5 + 2z - 1 = 0.$$

2. Obtain the values of $\sqrt[3]{1 + 2i}$.

3. Find the smallest roots (except $w = 0$) of $\sinh w = w$.

Hint: After separating the real and imaginary parts, plot $v = \cos^{-1}(u/\sinh u)$ and $u = \cosh^{-1}(v/\sin v)$.

Ans. $w = \pm 2.7687 \pm i7.4977$.

4. *Perturbation methods*

"Perturbation methods" depend on the use of one approximation to obtain a better one. The given equation is approximated by a simpler equation; the latter is solved and the roots are used to obtain successive corrections.

For example, if x is large, (6) may be replaced by

$$\tan x = \infty.$$

Solutions of this equation are

$$x_n = (n + \tfrac{1}{2})\pi, \quad n = 0, \quad \pm 1, \quad \pm 2, \cdots.$$

Let

$$x = (n + \tfrac{1}{2})\pi + \Delta_n \tag{22}$$

and substitute in (6) to obtain

$$\cot \Delta_n = -(n + \tfrac{1}{2})\pi - \Delta_n. \tag{23}$$

If Δ_n is small, $\cot \Delta_n \simeq 1/\Delta_n$ and

$$\Delta_n \simeq -2/(2n + 1)\pi.$$

Thus

$$x_n \simeq \left(n + \frac{1}{2}\right)\pi - \frac{2}{(2n + 1)\pi}.$$

These are the approximate roots when n is large.

For more accurate computation by successive approximations rewrite (23) in the form

$$\Delta_n = -\cot^{-1}[(n + \tfrac{1}{2})\pi + \Delta_n]. \tag{24}$$

Calculate Δ_n on the assumption that on the right side $\Delta_n = 0$; substitute this value in the bracketed expression and recalculate Δ_n; continue this process until the desired accuracy has been secured. Thus for $n = 1$ the sequence of successive approximations is

$$-0.209105, \qquad -0.218510, \qquad -0.218957, \qquad -0.218978, \cdots.$$

Hence the first positive root is $x = 4.49341$.

If we start with $n = 0$, we obtain the following sequence

-0.5669, -0.7834, -0.9038, -0.9825, -1.039, -1.082, -1.116, -1.144.

The limit of the sequence is $-\pi/2$ and the corresponding root from (22) is $x = 0$. This root is evident by inspection. The reason for slow convergence of this sequence is the extreme poorness of $\tan x = \infty$ as an approximating equation to $\tan x = x$ when $x = 0$; it is a wonder that the sequence converges at all.

The method is not limited to real roots; thus equation (16) may be written as

$$\tanh w = -ik/w. \tag{25}$$

If the absolute value of w is large compared with k,

$$\tanh w \simeq 0, \qquad w \simeq in\pi.$$

Let $w = in\pi + \Delta$ and substitute in (25); then

$$\tanh \Delta = -\frac{ik}{in\pi + \Delta}\cdot$$

For small Δ, $\tanh \Delta \simeq \Delta$ and $\Delta \simeq -k/n\pi$; hence $w \simeq in\pi - k/n\pi$.

On the other hand, if k is large compared with $|w|$, $\tanh w \simeq \infty$ and $w \simeq i(2n + 1)\pi/2$. Proceeding as above, we obtain

$$w \simeq \frac{(2n + 1)\pi}{2}\left(i - \frac{1}{k}\right)\cdot$$

Next consider a more general case

$$\frac{N(w)}{D(w)} = k, \tag{26}$$

where k is small.* Suppose that we know the zeros of the numerator, $N(\bar{w}) = 0$ We write $w = \bar{w} + \Delta$. By Taylor's theorem

$$\begin{aligned} N(w) &= N(\bar{w}) + N'(\bar{w})\Delta + \tfrac{1}{2}N''(\bar{w})\Delta^2 + \cdots \\ D(w) &= D(\bar{w}) + D'(\bar{w})\Delta + \tfrac{1}{2}D''(\bar{w})\Delta^2 + \cdots. \end{aligned} \tag{27}$$

Several cases may present themselves. If $D(\bar{w}) \neq 0$, $N'(\bar{w}) \neq 0$,

$$\Delta \simeq k\frac{D(\bar{w})}{N'(\bar{w})}, \qquad w \simeq \bar{w} + k\frac{D(\bar{w})}{N'(\bar{w})}\cdot \tag{28}$$

If $D(\bar{w}) \neq 0$, $N'(\bar{w}) = 0$,

$$\Delta \simeq \pm\sqrt{\frac{2kD(\bar{w})}{N''(\bar{w})}}, \qquad w \simeq \bar{w} \pm \sqrt{\frac{2kD(\bar{w})}{N''(\bar{w})}}\cdot \tag{29}$$

* If k is large, take the reciprocal of the equation.

If $D(\bar{w}) = N'(\bar{w}) = 0$ but $D'(\bar{w}) \neq 0$, $N''(\bar{w}) \neq 0$,

$$\Delta \simeq \frac{2kD'(\bar{w})}{N''(\bar{w})}. \tag{30}$$

In this manner we can find Δ in any eventuality, provided k is "sufficiently small."

By "sufficiently small" is meant that the absolute value of the correction Δ is small compared with the first approximation $\bar{w}$ to a root of $N(w) = 0$. This is the only sense in which we can describe k as "small"; for no matter how small k may be in equation (26) we can make the right side arbitrarily large by multiplying the equation by a large number so that the new "k" will not seem small. Conversely, the equation can be multiplied by a very small factor and the right side will then seem small.

Problems

1. Apply the perturbation method to find the large roots of (8).

2. Consider $\sinh w = w$ and show that, if w_0 is an approximate solution of the equation, $w - w_0 \simeq \dfrac{w_0 - \sinh w_0}{\cosh w_0 - 1}$. Using the answer to Problem 3 of Section 3, show that $w - w_0 = -0.00002172 - i0.00002372$.

3. Apply the perturbation method to $x^3 + 0.1x - 1 = 0$.
Hint: Let $x = \cos \varphi_0 + i \sin \varphi_0 + \Delta$, where $\varphi_0 = 0, \pm 2\pi/3$.

4. Apply the perturbation method to $ax^2 + bx + c = 0$ when b is large compared with a and c.
If the equation is transformed into $x = -(c/b) - (a/b)x^2$, the first approximation taken as $x_0 = -(c/b)$, and the following approximations obtained by successive substitutions, then only one root is found. Why? What are the relative magnitudes of this root and the one that remains? Having answered these questions obtain the second root in a similar form.

5. Obtain approximations to all three roots of $ax^3 + bx + c = 0$ when b is large compared with a and c.

6. Let z_0 be an approximate root of $f(z) = 0$. Show that $z = z_0 - [f(z_0)/f'(z_0)]$ is a better approximation if the correction term is small.

References

The methods described in this chapter have been chosen because they are applicable to any equation; they are not always the best for any particular equation or class of equations, for they do not take advantage of special properties of functions entering the equation. If the reader is only occasionally concerned with the solution of equations, it may require more effort on his part to learn special methods than to obtain his answers directly as explained in this chapter; this is particularly true if the equations encountered by him are of varied forms.

A paper by H. C. Plummer, *Philosophical Magazine*, December, 1941, pp. 505–512, will introduce the reader to the idea of taking advantage of certain properties of circular functions for obtaining rapidly accurate values of the roots of certain equations containing them.

Most special methods have been developed for algebraic equations. A good method is given by Shih-Nge Lin in the *Journal of Mathematics and Physics*, Vol. 20, No. 3, August, 1941, pp. 231–242. In a paper by Thornton C. Fry in the *Quarterly of Applied Mathematics*, July, 1945, pp. 89–105, the reader will find a discussion of some of these methods as well as valuable references. Pages 2–15 of the " Smithsonian Mathematical Formulae and Tables of Elliptic Functions " deal with some of the more classical methods.

CHAPTER IV

Power Series

1. *Arithmetic series*

Infinite series are first encountered in arithmetic. Infinite decimals $1.333\cdots$, $1.4142\cdots$, are series of constant terms; thus

$$1.333\cdots = 1 + \tfrac{3}{10} + \tfrac{3}{100} + \tfrac{3}{1000} + \cdots,$$
$$1.4142\cdots = 1 + \tfrac{4}{10} + \tfrac{1}{100} + \tfrac{4}{1000} + \cdots,$$

more generally,

$$S = a_1 + a_2 + a_3 + \cdots + a_n + \cdots. \tag{1}$$

We could not get along without such series. Although it is true that the first of the above series represents $\frac{4}{3}$ and hence could be dispensed with, there is nothing that we can do about the second which represents the square root of two. To denote this number by $\sqrt{2}$ is not to give its value; to "know" $\sqrt{2}$ is to know a process for finding a sequence of common fractions whose squares approach the limit 2. Infinite sequences of numbers $x_1, x_2, x_3, \cdots$ and series $x_1 + x_2 + x_3 + \cdots$ are the foundation of general arithmetic. Irrational numbers are first *defined* as limits of sequences or sums of series of rational numbers; then, irrational numbers themselves may be taken as elements of sequences or series; and the arithmetic structure rises.

An infinite series is said to be *convergent* and its sum is said to be S if $|S - S_n|$, where

$$S_n = a_1 + a_2 + \cdots + a_n, \tag{2}$$

can be made less than any preassigned small quantity ϵ for all values of n greater than some value N. It is evident that the series will be *divergent* (not convergent) unless a_n approaches zero as n increases indefinitely.

The requirement that a_n should approach zero is a necessary but not a sufficient condition for convergence. Thus

$$S = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \tag{3}$$

is divergent. In order to prove this, write

$$S = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)$$
$$+ \left(\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16}\right) + \cdots.$$

The numbers in parentheses are all greater than $\frac{1}{2}$ and the "partial sum" S_n can be made as large as we please by taking n large enough.

Problems

1. Start with a pair of numbers a, b. Form their arithmetic mean $a_1 = \frac{1}{2}(a + b)$ and "harmonic mean" $b_1 = 2ab/(a + b)$ obtained from $\frac{1}{b_1} = \frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right)$. From a_1, b_1 obtain a_2, b_2 as above, etc. Show that the sequences $a, a_1, a_2, \cdots$ and $b, b_1, b_2, \cdots$ converge to the same limit and that this limit is the geometric mean $\sqrt{ab}$. Apply this method to the computation of $\sqrt{2}$.

2. Show that the positive solution of the equation $x^2 = 2$ may be represented as a "continued fraction"

$$x = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}\,.$$

Calculate the successive convergents of this continued fraction by terminating the fraction.

Hint: Let $x = 1 + y$ and "solve" for y in a form suitable to the above representation.

3. Show that the positive solution of $x^2 = a$, where a lies between the squares, n^2 and $(n + 1)^2$, of two successive integers may be expressed as

$$x = n + \cfrac{a - n^2}{2n + \cfrac{a - n^2}{2n + \cdots}}\,.$$

4. Generalize the method in Problem 1 so that it can be applied to the computation of the nth root of a number.

2. *Absolute and relative convergence*

The sum of a finite number of terms, such as (2), is independent of the order of addition. This is not necessarily true of an infinite series. Naturally, if we reshuffle the terms anywhere in the finite part of the series the sum will not be affected; but the sum may be altered if the reshuffling is done on an infinite scale.

Consider, for instance,

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots. \tag{4}$$

The sum of this series *in the order indicated* is known to be log 2 (the natural

logarithm). Suppose we derange the terms as follows

$$\hat{S} = 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots. \tag{5}$$

In the new series we have included all terms of (4) and added none; we are merely adding two positive terms for each negative. It can be shown that $\hat{S} = 1.5 \log 2$. We write

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}. \tag{6}$$

Then

$$\tfrac{1}{2}s_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \cdots + \frac{1}{2n}$$

$$s_{2n} - \tfrac{1}{2}s_n = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1},$$

$$S_{2n} = s_{2n} - s_n, \tag{7}$$

$$\begin{aligned}\hat{S}_{3n} &= \left(1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots + \frac{1}{4n-1}\right) - \left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\right)\\ &= (s_{4n} - \tfrac{1}{2}s_{2n}) - \tfrac{1}{2}s_n = (s_{4n} - s_{2n}) + \tfrac{1}{2}(s_{2n} - s_n)\\ &= S_{4n} + \tfrac{1}{2}S_{2n}.\end{aligned}$$

As $n \to \infty$, we have

$$\hat{S} = S + \tfrac{1}{2}S = 1.5S. \tag{8}$$

Series whose sums are independent of the order of summation are called *absolutely convergent;* other convergent series are only *relatively convergent.*

The reader may already suspect that relative convergence is connected with the presence of positive and negative terms, and with the divergence of parts of the complete series made up of the terms of one sign. This is actually the case. If the series of absolute values

$$\bar{S} = |a_1| + |a_2| + |a_3| + \cdots + |a_n| + \cdots \tag{9}$$

is convergent at all it is absolutely convergent and so is the original series (1). On the other hand, if the positive and negative parts of the series diverge when taken separately, the value of the sum of the series will be altered by any rearrangement of the positive and negative terms. If a large positive sum is desired, more positive terms are taken for every negative term.

It may be suggested that to tamper with the order of the terms is to create an artificial situation and that in practice we need not worry about the distinction between absolute and relative convergence. But the

distinction is important if we multiply two series,

$$\begin{aligned} A &= a_1 + a_2 + a_3 + \cdots + a_n + \cdots, \\ B &= b_1 + b_2 + b_3 + \cdots + b_n + \cdots, \end{aligned} \tag{10}$$

for the order of summation of the product series is not given naturally. If the two series are absolutely convergent, we can multiply them term by term in any order, the resulting series will be absolutely convergent and its sum will be the product of the sums of the original series. If A and B are relatively convergent, we should be careful about the order of the product series.

It is fair to add, however, that only slowly convergent series may be relatively convergent and that care has to be exercised only in borderline cases.

3. *Tests of convergence*

There are numerous tests of convergence; but from the practical point of view the most important is the *comparison test*. Let

$$\begin{aligned} S &= a_1 + a_2 + a_3 + \cdots + a_n + \cdots \\ T &= b_1 + b_2 + b_3 + \cdots + b_n + \cdots \end{aligned} \tag{11}$$

be two series of positive numbers, and suppose that the first series is known to be convergent; then, if " almost all " (meaning " all except a finite number of ") terms of the second series are not larger than the corresponding terms of the first series, the second series is also convergent. The theorem is so obvious intuitively that we shall dispense with a formal proof.

Similarly if the first series is divergent and " almost all " terms of the second series are not smaller than the corresponding terms of the first series, then the second series is also divergent.

Note that in proving the divergence of (3) we actually employed the comparison test. Since the series of absolute values of any series is a series of positive numbers, the above test is a test of the absolute convergence of any series.

Problems

1. Find a simple direct argument proving that the series

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

is convergent and that its sum is 2.

2. Using a method analogous to that in Section 1, show that the following series

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

is convergent.

3. Prove that

$$1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \cdots$$

is convergent.

4. Prove that

$$1 + \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{4}\log 4 + \cdots$$

is divergent.

5. Prove that

$$\frac{1}{1\cdot 2\cdot 3} + \frac{1}{4\cdot 5\cdot 6} + \frac{1}{7\cdot 8\cdot 9} + \cdots$$

is convergent.

4. *Method of increasing the rapidity of convergence of infinite series*

Several methods of increasing the rapidity of convergence of infinite series are explained in that useful collection, "Smithsonian Mathematical Formulae," on pages 113–115. We shall not consider these methods but shall confine our attention to the most obvious one

Consider two series and their difference

$$\begin{aligned} U &= u_0 + u_1 + u_2 + u_3 + \cdots \\ V &= v_0 + v_1 + v_2 + v_3 + \cdots \\ U - V &= (u_0 - v_0) + (u_1 - v_1) + (u_2 - v_2) + \cdots. \end{aligned} \tag{12}$$

If V is known or can be found and if the difference series converges more rapidly than the u-series, we have, in effect, a transformation increasing the rapidity of convergence of the u-series. For example, consider

$$\begin{aligned} U &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots \\ &= \frac{1}{1\cdot 2} + \frac{1}{3\cdot 4} + \frac{1}{5\cdot 6} + \cdots = \sum_{n=1}^{\infty} \frac{1}{(2n-1)2n}. \end{aligned} \tag{13}$$

It is known that

$$S = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \tag{14}$$

the difference,

$$U - \tfrac{1}{4}S = \frac{1}{4}\sum_{n=1}^{\infty} \frac{1}{n^2(2n-1)}, \tag{15}$$

converges more rapidly than (13).

Take as another example,

$$S = \sum_{n=1}^{\infty} \frac{\cos nx}{n+1}. \tag{16}$$

On page 138 of the Smithsonian Tables we find

$$\begin{aligned} U &= \sum_{n=1}^{\infty} \frac{\cos nx}{n} = \frac{1}{2}\log\frac{1}{2(1-\cos x)}, \quad 0 < x < 2\pi; \\ V &= \sum_{n=1}^{\infty} \frac{\cos nx}{n^2} = \frac{\pi^2}{6} - \frac{\pi x}{2} + \frac{x^2}{4}, \quad 0 < x < 2\pi. \end{aligned} \tag{17}$$

Subtracting U from S and adding V, we obtain

$$S = U - V + \sum_{n=1}^{\infty} \frac{\cos nx}{n^2(n+1)}. \tag{18}$$

The series on the right converges more rapidly than the original series.

This example has been chosen merely to illustrate the method; actually S can be expressed in terms of series whose sums are known, without any residue series. Thus, introducing $m = n + 1$, we have

$$\begin{aligned} S &= \sum_{m=2}^{\infty} \frac{\cos(m-1)x}{m} = \cos x \sum_{m=2}^{\infty} \frac{\cos mx}{m} + \sin x \sum_{m=2}^{\infty} \frac{\sin mx}{m} \\ &= (U - \cos x)\cos x + \left(\frac{\pi - x}{2} - \sin x\right)\sin x \\ &= \frac{\pi - x}{2}\sin x + \frac{1}{2}\cos x \log\frac{1}{2(1-\cos x)} - 1. \end{aligned} \tag{19}$$

This method of summation will not work if the denominator in (16) is $n + 0.5$, or $n + k$, where k is a fraction.

5. *Power series*

The simplest and most direct arithmetic definition of a function of a single variable is in terms of power series; and this definition is broad enough to apply to complex values of the independent variable. The function $w = \sin z$ has a very simple geometric definition if z is real; but the definition fails if z is complex. On the other hand, a definition by a series of powers of z,

$$\sin z = z - \frac{1}{6}z^3 + \frac{1}{120}z^5 - \frac{1}{7!}z^7 + \frac{1}{9!}z^9 - \cdots,$$

works equally well for either real or complex values of z. We have to know only how to add, subtract and multiply numbers; and that is all.

On the other hand, if $\rho > 1$, the remainder term increases as n increases and the series diverges. If $\rho = 1$, the remainder term is finite as long as $\text{ph}\, z \neq 0$, but it does not approach zero; the amplitude of z^n is always unity but the phase changes with n; hence the series is divergent.

A more general geometric series is obtained from (20) if we replace z by z/A, where $A = a(\cos\varphi + i\sin\varphi)$ is some constant; thus

$$S = 1 + (z/A) + (z/A)^2 + (z/A)^3 + \cdots. \tag{23}$$

This series converges when $\rho < a$, and diverges when $\rho \geq a$.

Geometric series are convenient for comparison purposes in proving either convergence or divergence of series.

Problems

1. Prove that

$$S = \sin x + 2\sin\frac{x}{5} + 4\sin\frac{x}{25} + \cdots + 2^n \sin\frac{x}{5^n} + \cdots$$

converges for all values of x.

2. Prove that a series of constant terms, $a_0 + a_1 + a_2 + \cdots + a_n + \cdots$, converges if $\lim |a_{n+1}/a_n|$ exists and is less than unity. This is the *ratio test.*

3. Give an example in which the limit mentioned in Problem 2 does not exist and yet the series is convergent (even rapidly convergent).

7. *Circle of convergence*

In the complex plane the geometric series (23) converges everywhere inside the circle of radius a with its center at the origin and diverges everywhere outside the circle and on its circumference. With only a slight modification this property is shared by all power series. There is always a circle inside which a given power series converges, while diverging outside. On the circumference itself it may either converge or diverge. This is the *circle of convergence.* Its radius may be equal to zero, so that the series would diverge everywhere in the plane except at $z = 0$; or it may be infinite and the series would converge everywhere.

The assertion is almost self-evident. Taking a power series

$$S = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots \tag{24}$$

we assume at first that the coefficients are positive real. The series of absolute values is then

$$\bar{S} = a_0 + a_1\rho + a_2\rho^2 + \cdots + a_n\rho^n + \cdots. \tag{25}$$

If this series converges for some particular value $\rho = \rho_0$, it converges for all smaller values (the comparison test). When z is complex, there are in

effect two power series

$$\begin{aligned} S' &= a_0 + a_1\rho\cos\varphi + a_2\rho^2\cos 2\varphi + \cdots, \\ S'' &= a_1\rho\sin\varphi + a_2\rho^2\sin 2\varphi + \cdots. \end{aligned} \tag{26}$$

By the comparison test these series are convergent for $\rho \leq \rho_0$; that is, inside the circle of radius ρ_0 and on its circumference.

Let us keep increasing ρ. If (25) remains convergent for all values of ρ, (24) will be convergent in the entire plane. If for some $\rho = a$ or for some ρ just exceeding a, (25) fails to converge, it will fail for all larger values as well. The following series illustrate the distinction between the two possibilities

$$\begin{aligned} S &= 1 + \rho + \frac{1}{2}\rho^2 + \frac{1}{3}\rho^3 + \cdots + \frac{1}{n}\rho^n + \cdots, \\ T &= 1 + \rho + \frac{1}{4}\rho^2 + \frac{1}{9}\rho^3 + \cdots + \frac{1}{n^2}\rho^n + \cdots. \end{aligned} \tag{27}$$

The first of these series stops converging as soon as ρ reaches unity and the second as soon as ρ exceeds unity. Normally, the divergence of the series is evident for the nth term grows larger and larger; in this case, divergence of (26) and hence of (24) is also assured. If we can establish the fact that divergence of (25), coupled with the property that the nth term approaches zero, can take place for only one value of ρ, we can narrow down the region of uncertainty about the behavior of (24) to the circumference of a circle. So let us suppose that (25) diverges when $\rho = a$ and yet the nth term $\epsilon_n = a_n a^n$ approaches zero with increasing n. For a slightly smaller value $\rho = a - \delta$ the nth term is

$$a_n(a - \delta)^n = a_n a^n \left(1 - \frac{\delta}{a}\right)^n = \epsilon_n \left(1 - \frac{\delta}{a}\right)^n.$$

Since ϵ_n becomes eventually smaller than unity, this nth term is smaller than the corresponding term of a convergent geometric series. Hence the series converges if $\rho < a$.

Suppose now that $\rho = a + \delta$; the nth term is then

$$\eta_n = \epsilon_n \left(1 + \frac{\delta}{a}\right)^n.$$

Although the second factor increases with n, the first decreases and η_n might yet approach zero; but if it did, ϵ_n would eventually be smaller than the nth term $[1 + (\delta/a)]^{-n}$ of a convergent geometric series, and consequently the power series would converge when $\rho = a$, which is contrary to our initial assumption. Thus the uncertainty about the convergence of (24) is restricted to the circumference of a circle. There anything can

happen: the series may diverge everywhere as in (20); it may diverge at one point and converge at others as in the general series corresponding to S in (27); and the series may converge over the entire circumference as in the general power series corresponding to T in (27).

The theorem has been established for the case of positive real coefficients. With only minor modifications the argument applies to any power series. The a_n's in (25) will be the amplitudes of the corresponding coefficients in the general series; and in (26) the phases of the coefficients will be added to $n\varphi$; but nothing essential in the argument will be affected.

A series of negative powers

$$S = \frac{b_1}{z} + \frac{b_2}{z^2} + \cdots + \frac{b_n}{z^n} + \cdots \tag{28}$$

is converted into a series of positive powers by the substitution

$$z = 1/w. \tag{29}$$

This transformation exchanges the inner and outer regions of the circle of convergence so that (28) will converge *outside* its circle of convergence and diverge inside.

A mixed power series, if convergent, converges inside a *ring of convergence*, Figure 4.1. The inner boundary must be the boundary for the circle of convergence of the series of negative powers and the outer boundary must be the boundary of convergence for the series of positive powers. It is quite possible for the ring of convergence to be reduced to just one circumference or even cease to exist. The outer boundary of a ring may very well be at infinity; the inner boundary may shrink to a point, $z = 0$. The series (28) can never converge at $z = 0$. The slight asymmetry in the behavior of power series disappears if z is represented on the surface of the Neumann sphere; there the outer boundary also shrinks to a point, the point at infinity.

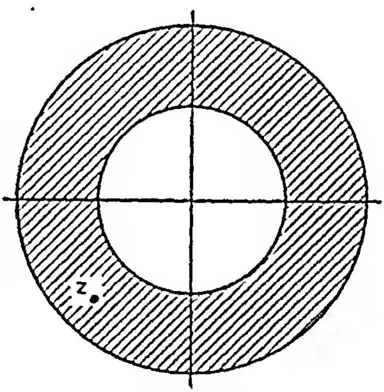

FIG. 4.1. In general, a series which contains positive and negative integral powers of z converges only inside a circular *ring of convergence*.

8. *Radius of convergence*

Everything points strongly to all power series being almost geometric series (23) or sums of such series. Series (24) can be represented as

$$S = a_0 + a_1 z + (a_2^{1/2} z)^2 + (a_3^{1/3} z)^3 + \cdots + (a_n^{1/n} z)^n + \cdots. \tag{30}$$

If the sequence $|a_n|^{1/n}$ converges to a limit a, $1/a$ is the radius of convergence. If it does not converge it may have an *upper limit* a; then $1/a$ is the radius of convergence. It is perfectly possible for a sequence of numbers to fail to converge simply because it consists of several sequences converging to different limits; it is the reciprocal of the largest of these limits that is equal to the radius of convergence. The sum of two series with different circles of convergence will converge only inside the smaller circle.

The circle of convergence has an important bearing on the possibility of using power series to represent *physical* functions. In such functions the variable is real and a particular function may be well behaved in a given interval and yet the power series representing the function may diverge in this interval. The physical existence of the function is no indication that its power series will converge; this is because the behavior of the power series is governed by the behavior of the function in the complex plane even though complex values may be irrelevant in the particular physical problem. For example, the power series for $f(z) = 1/(1 + z^2)$ will diverge for real values of z equal to or greater than unity simply because it has to diverge when $z = i$. If $z = i$, $z^2 = -1$ and $f(i) = \infty$; if the series is to represent the function, it must diverge for $z = i$ for this is the only way for the series to assume an infinite value. For purely mathematical reasons, this divergence at one point requires the divergence of the series everywhere outside the circle centered at the origin and passing through $z = i$.

Problem. Show that if $\lim |a_n/a_{n+1}| = a$, then a is the radius of convergence. More generally if a is the lower limit of the ratio, a is the radius of convergence.

9. *Uniform convergence*

Uniform convergence refers only to series of functions. Consider

$$S(x) = 1 + x + x^2 + \cdots + x^n + \cdots \tag{31}$$

whose nth partial sum is

$$S_n(x) = \frac{1}{1-x} - \frac{x^n}{1-x}. \tag{32}$$

The series is absolutely convergent everywhere inside the interval $(-1,1)$, for we can always choose n large enough to make x^n and hence the remainder $S(x) - S_n(x)$ as small as we please. However, as x approaches unity, n has to be chosen larger and larger in order that the remainder may be made smaller than a preassigned small quantity.

A series is said to be *uniformly convergent in a given region* if, corresponding to a given small positive value ϵ, there is a positive number N *independent of the variable* such that for all $n > N$ the absolute value of the remainder is less than ϵ everywhere in the region. Series (31) is not uni-

formly convergent in the interval $(-1,1)$ but it is uniformly convergent in the interval $(-1,1-\delta)$, where δ is some positive quantity, no matter how small. To show this we decide on ϵ and find N for $x = 1 - \delta$; the same N will do for all x in the interval. Nonuniform convergence is a property of the neighborhoods of certain exceptional points and, like relative convergence, it is associated with slow convergence.

The idea of uniform convergence is essential in extending certain properties of finite sums to infinite series. Thus, a finite sum of continuous functions is continuous; but the sum of an infinite series of continuous functions is not necessarily continuous. The sum is continuous if the series is uniformly convergent. A spectacular example of discontinuity is furnished by the following series

$$S(x) = x^2 + \frac{x^2}{1+x^2} + \frac{x^2}{(1+x^2)^2} + \cdots.$$

This series is the product of x^2 and a series just like (31) with x replaced by $1/(1+x^2)$; hence the nth partial sum is

$$S_n(x) = 1 + x^2 - \frac{1}{(1+x^2)^{n-1}}.$$

As $n \to \infty$, we have

$$S(x) = \lim S_n(x) = 1 + x^2$$

for all x except $x = 0$. As $x \to 0$, $S(x)$ approaches unity as its limit; but if $x = 0$, all partial sums vanish and the sum of the series is zero. The series happens to be nonuniformly convergent in the vicinity of $x = 0$; also it is a slowly convergent series in this vicinity. In fact, $x^{-2}S(x)$ diverges at $x = 0$ and $S(x)$ converges only on account of the factor* x^2.

10. *Differentiation and integration of power series*

Power series obtained by differentiating and integrating a power series term by term have the same radius of convergence as the original series; they represent the derivative and integral of the sum

$$\begin{aligned} dS/dz &= a_1 + 2a_2z + 3a_3z^2 + \cdots + na_nz^{n-1} + \cdots, \\ \int_a^z S(z)\,dz &= a_0(z-a) + \frac{1}{2}a_1(z^2-a^2) + \frac{1}{3}a_2(z^3-a^3) + \cdots \\ &= \left(a_0z + \frac{1}{2}a_1z^2 + \frac{1}{3}a_2z^3 + \cdots + \frac{1}{n}a_nz^{n+1} + \cdots\right) \\ &\quad - \left(a_0a + \frac{1}{2}a_1a^2 + \frac{1}{3}a_2a^3 + \cdots + \frac{1}{n}a_na^{n+1} + \cdots\right). \end{aligned} \tag{33}$$

* For other examples of nonuniform convergence and graphical illustrations, see Ivan S. Sokolnikoff, *Advanced Calculus*, McGraw-Hill Book Company, Inc., New York, 1939, pp. 253–255; also the Weierstrass test on p. 262.

The nth term of the derivative series is $a_n[zn^{1/(n-1)}]^{n-1}$; if z is inside the circle of convergence of the original series, then for a sufficiently large N the quantity $zN^{1/(N-1)}$ is also inside the circle. Replacing every n by N, we obtain a convergent series which dominates the derivative series; thus the convergence of the latter has been proved.

This proof does not apply to a point z on the circumference itself; then $zN^{1/(N-1)}$ is always outside the circle of convergence. It is clear that the derivative series converges more slowly than the original series.* If the latter is divergent, the former is also divergent; if the latter is convergent, the former may be either convergent or divergent.

The integral series converges more rapidly than the original series. If the latter converges on the circumference, the former also converges; if the latter diverges, the former may either diverge or converge.

In obtaining (33), differentiation and integration have been performed as if z were real; in the next two chapters it is shown that this is permissible.

We have still to prove that the two series in (33) actually represent the derivative and integral of $S(z)$. The identification may seem natural; but there is a question concerning the interchange of limiting processes. The derivative and integral are obtained as limits; the sum of a series is obtained as a limit. It is one thing to integrate each term of an infinite series and then add; it is something else to add the series and then integrate the sum. It is easier to believe that this interchange of operations is permissible for finite sums; and a simple proof may be found in texts on elementary calculus. This proof can be generalized if we separate the original series into the first n terms and the remainder $R_n(z)$ and thus reduce it to a finite sum. All that we have to prove then is that the integral of the remainder approaches zero as n increases indefinitely. Here is where the conception of uniform convergence enters. For the present purpose it is not enough to be assured that, for a sufficiently large n, the remainder can be made less than some arbitrarily small quantity ϵ for any particular z; $|R_n(z)|$ should be less than ϵ for *all z in the interval of integration* when n exceeds a certain value N. It is easy to show that power series are uniformly convergent in the interior of the circle of convergence; the truth of the second equation in (33) then follows. The same proof applies to the integral of the first equation in (33); and this establishes it automatically.

Thus we have found once more that the " natural inclination " to treat infinite series as if they were finite sums may be followed safely when the series converge rapidly; but legitimate doubts arise in borderline cases and there are examples showing that these doubts are well founded. In con-

* That is, when we compare the corresponding terms of the two series.

clusion, we should add that from time to time such borderline cases actually arise even in applied mathematics, sometimes because of idealization of natural phenomena and sometimes because of inherent properties of a particular mathematical representation of such phenomena.

Problems

1. Suppose that our knowledge of functions is limited to power functions x^n; and to their sums, differences, products, and quotients. The integral

$$f(x) = \int_1^x x^n\,dx = \frac{1}{n+1}(x^{n+1} - 1) \tag{A}$$

is well defined if $n \neq -1$. If $n = -1$, the right side does not exist. Does this mean that

$$F(x) = \int_1^x \frac{dt}{t} \tag{B}$$

does not exist? Of course, the answer is "no." There are two possible reasons for the failure:

1. $F(x)$ does not exist;
2. the particular mathematical form of the answer in (A) is not suitable for the representation of $F(x)$.

The second reason is the answer in the present case.

Calculate $F(x)$ by first expanding $1/t$ in a series of powers of $(1 - t)$ and then integrating term by term. What is the region of convergence of the series?

Ans. $-(1 - x) - \frac{1}{2}(1 - x)^2 - \frac{1}{3}(1 - x)^3 - \cdots$; the circle centered at $x = 1$ and passing through $x = 0$.

Note: $F(x)$ has a name: the *natural logarithm* of x.

2. With the aid of power series obtain

$$f_1(x) = \int_0^x \frac{dx}{1 + x^2}, \qquad f_2(x) = \int_0^x \frac{dx}{1 - x^2}, \qquad f_3(x) = \int_0^x \frac{dx}{1 + x^n}.$$

Ans.
$$f_1(x) = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots,$$
$$f_2(x) = x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \frac{1}{7}x^7 + \cdots,$$
$$f_3(x) = x - \frac{1}{n+1}x^{n+1} + \frac{1}{2n+1}x^{2n+1} - \frac{1}{3n+1}x^{3n+1} + \cdots.$$

Note: $f_1(x)$ and $f_2(x)$ have names: the *inverse circular tangent* of x and the *inverse hyperbolic tangent* of x.

11. *Taylor's series*

By successive differentiation of the power series (24) we find that, at $z = 0$,

$$\begin{gathered} S(0) = a_0, \quad S'(0) = a_1, \quad S''(0) = 1 \cdot 2a_2, \cdots \\ S^{(n)}(0) = 1 \cdot 2 \cdot 3 \cdots na_n, \cdots \end{gathered} \tag{34}$$

and therefore

$$S(z) = S(0) + S'(0)z + \frac{1}{2!} S''(0)z^2 + \cdots + \frac{S^{(n)}(0)}{n!} z^n + \cdots. \quad (35)$$

This is *Maclaurin's formula* in the complex plane.

More generally, we can write a series of powers of $(z - a)$ which will converge inside some circle centered at $z = a$; for the sum of this series we have

$$f(z) = f(a) + f'(a)(z - a) + \cdots + f^{(n)}(a) \frac{(z - a)^n}{n!} + \cdots. \quad (36)$$

This is *Taylor's formula* in the complex plane.

Both these series are treated in elementary calculus but only for real variables. This restriction sometimes makes it difficult to understand their behavior. Take the following classical example of a function and its derivatives

$$\begin{aligned} f(x) &= e^{-1/x^2}, \qquad f'(x) = (2/x^3)e^{-1/x^2}, \\ f''(x) &= -(6/x^4)e^{-1/x^2} + (4/x^6)e^{-1/x^2} \cdots . \end{aligned} \quad (37)$$

The function and all its derivatives vanish at $x = 0$ and Maclaurin's series reduces identically to zero. Now if the series were divergent for all values of x, we might be disappointed but not perturbed; we should merely say that the function under consideration could not be represented by a power series of this particular form. As it is, the situation is distressing and, in the domain of real variables, nothing can be done to remove the uncertainty except by writing Maclaurin's and Taylor's formulas as *finite series with remainder terms.* The remainder terms depend on the values of the derivatives at points *outside* the particular point for which the coefficients of the power series are being calculated. For instance, for the above function Maclaurin's formula should be written as

$$e^{-1/x^2} = 0 + 0 \cdot x + 0 \cdot x^2 + \cdots + 0 \cdot x^{n-1} + \frac{x^n}{n!} f^n(\xi) \quad (38)$$

where ξ is some quantity larger than zero. More specifically, the finite series (38) is valid in any interval $0 \leq x \leq a$, so long as ξ is within that interval. Since ξ is unknown and the remainder term constitutes the entire significant part of the series, we cannot derive any benefit from the expansion.

No such difficulties ever arise if we consider power series in the complex plane. Replacing x by the complex variable z, and studying the values of $f(z)$ and its derivatives at $z = 0$, we find that none of them exist there. For if $z = iy$, $f(z) = \exp(1/y^2)$ and this increases indefinitely as y ap-

proaches zero. Hence, regardless of the value we might assign to the function itself at $z = 0$, its derivative does not exist there and that puts an end to the possibility of representing the function by Maclaurin's series. This may be disappointing; but at least we know where we stand without a difficult investigation. If we have a function such that, at a certain point, the function and all its derivatives exist *in the complex plane*, then the function can be represented by a power series which will converge at least within *some* circle centered at this point.

Problems

1. Expand $(1 - x)^{-1/2}$ in a power series.

Ans. $1 + \frac{1}{2}x + \frac{3}{8}x^2 + \frac{5}{16}x^3 + \cdots \frac{(1/2)(3/2)(5/2)\cdots[(2n-1)/2]}{1\cdot 2\cdot 3\cdots n}x^n + \cdots.$

2. Expand $(1 - x^2)^{-1/2}$ in a power series.

3. Calculate $f(x) = \int_0^x \frac{dx}{\sqrt{1-x^2}}$.

Ans. $f(x) = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \cdots + \frac{(1/2)(3/2)\cdots[(2n-1)/2]}{(2n+1)1\cdot 2\cdot 3\cdots n}x^{2n+1} + \cdots.$

Note: $f(x) = \sin^{-1} x$.

4. Calculate $f(x) = \int_0^x \sqrt{1-x^2}\,dx$.

5. The *elliptic integral of the first kind* is

$$F(x) = \int_0^x \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} = \int_0^\varphi \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}, 0 \le k \le 1,$$

where $x = \sin\varphi$. The *complete elliptic integral of the first kind*, K, corresponds to $x = 1, \varphi = \pi/2$.

Expand the second form in powers of $k \sin\varphi$. Using the formulas for the integrals of the powers of the sine function, show that

$$K = \frac{\pi}{2}\left\{1 + \sum_{n=1}^{\infty}\left[\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}\right]^2 k^{2n}\right\}.$$

Show that, if k is small, then approximately

$$F(x) \simeq \varphi + \tfrac{1}{4}k^2(\varphi - \tfrac{1}{2}\sin 2\varphi) = \sin^{-1} x + \tfrac{1}{4}k^2(\sin^{-1} x - x\sqrt{1-x^2}).$$

What is the next term in the expansion in powers of k?

6. The *elliptic integral of the second kind* is

$$E(x) = \int_0^x \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx = \int_0^\varphi \sqrt{1-k^2\sin^2\varphi}\,d\varphi, 0 \le k \le 1.$$

The *complete elliptic integral of the second kind*, E, corresponds to $x = 1, \varphi = \pi/2$.
Show that

$$E = \frac{\pi}{2}\left\{1 - \sum_{n=1}^{\infty}\left[\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n}\right]^2 \frac{k^{2n}}{2n-1}\right\}.$$

Find an approximate value of the incomplete integral when k is small.

7. By using power series show that

$$\lim \frac{\sqrt{1+x} - \sqrt{1-x}}{\sqrt{1+2x} - \sqrt{1-3x}} = \frac{2}{5} \qquad \text{as } x \to 0.$$

8. Show that

$$\lim \frac{\sqrt{1+x+x^2} - \sqrt{1+x-x^2}}{\sqrt{1+2x+3x^2} - \sqrt{1+2x+4x^2}} = -2 \qquad \text{as } x \to 0.$$

9. Show that

$$\lim \frac{\sqrt{3+2x} - \sqrt{3-x}}{\sqrt{5+x} - \sqrt{5+2x}} = -\sqrt{15} \qquad \text{as } x \to 0.$$

12. *Multiple-valued functions*

In Section 8 it has been pointed out that one perfectly good reason for the divergence of a power series in $(z - a)$ is that the function represented by it becomes infinite at some point z_0. The power series has to diverge at that point in order to yield an infinite value; on account of the nature of power series the divergence at one point forces divergence everywhere outside the circle centered at $z = a$ and passing through $z = z_0$.

The power series may diverge in "self-defense." Suppose we take the following function

$$w = (1 - z^2)^{1/2}. \tag{39}$$

It is a *multiple-valued function* since the square root of a number may be either positive or negative; but a series of positive integral powers of z, if it converges at all, represents automatically a single-valued function. Our first thought may be that the above function could not be represented by a power series anywhere; but this is not quite the case. We can certainly fix on one value of the function, $w = 1$, at $z = 0$. For small values of z we can fix on that particular value of the square root which approaches unity as z approaches zero. In this way we shall be able to calculate the values of the function and its derivatives at $z = 0$ for one "branch" of w; then we can apply Maclaurin's expansion. To simplify the process, we first expand a related function

$$f(t) = (1 + t)^{1/2}, \tag{40}$$

for which

$$\begin{aligned} f'(t) &= \tfrac{1}{2}(1+t)^{-1/2}, \quad f''(t) = \tfrac{1}{2}(-\tfrac{1}{2})(1+t)^{-3/2}, \\ f'''(t) &= \tfrac{1}{2}(-\tfrac{1}{2})(-\tfrac{3}{2})(1+t)^{-5/2} \cdots. \end{aligned} \tag{41}$$

The corresponding expansion is

$$f(t) = \pm(1 + \tfrac{1}{2}t - \tfrac{1}{8}t^2 + \tfrac{1}{16}t^3 - \tfrac{5}{128}t^4 + \cdots), \tag{42}$$

where the upper and lower signs correspond to $f(0) = 1$ and $f(0) = -1$. Substituting $t = -z^2$, we have

$$w = \pm(1 - \tfrac{1}{2}z^2 - \tfrac{1}{8}z^4 - \tfrac{1}{16}z^6 - \cdots). \tag{43}$$

Another more laborious method does not depend on differentiation. We may assume a series for w and square it:

$$\begin{aligned} w &= a_0 + a_1z^2 + a_2z^4 + a_3z^6 + \cdots \\ w^2 &= a_0^2 + 2a_0a_1z^2 + (a_1^2 + 2a_0a_2)z^4 + \cdots. \end{aligned} \tag{44}$$

Equating to $(1 - z^2)$, we obtain

$$a_0^2 = 1, \qquad 2a_0a_1 = -1, \qquad a_1^2 + 2a_0a_2 = 0 \cdots. \tag{45}$$

This sequence of equations in the coefficients enables us to find their values step by step. Only the first equation is ambiguous, leading to $a_0 = \pm 1$; if one of these values is chosen, the rest are uniquely determined.

Our series represent two single-valued functions; but, as we shall presently show, over the *entire* plane w is not just a pair of single-valued functions but *one* multiple-valued function. This leaves no alternative for series (43) but to diverge in some parts of the plane, and for power series this means to diverge outside some circle.

At points $z = \pm 1$, both values of w are the same (both equal to zero). The points at which two or more values of a multiple-valued function coalesce are called the *branch points*. It is around the branch points that we discover that w behaves as a single function and not as a collection of separate functions. Let

$$z = 1 + a(\cos\vartheta + i\sin\vartheta), \tag{46}$$

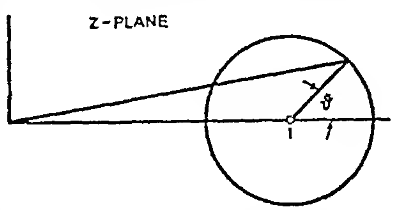

FIG. 4.2. As point z moves once round the circumference of a small circle centered at $z = 1$, the two values (48) of w defined by (39) are interchanged.

where a is small and constant while ϑ is allowed to vary. Graphically, this means that z will move round a circle centered at $z = 1$, Figure 4.2. Now $1 + z = 2 + a(\cos\vartheta + i\sin\vartheta)$ and $1 - z = -a(\cos\vartheta + i\sin\vartheta)$; for small a, $1 + z \simeq 2$ and

$$\begin{aligned} w = \sqrt{1 - z^2} &\simeq \sqrt{-2a(\cos\vartheta + i\sin\vartheta)} \\ &= \sqrt{2a}[\cos(\vartheta + \pi) + i\sin(\vartheta + \pi)]. \end{aligned} \tag{47}$$

Hence, the two values of w are approximately

$$\begin{aligned} w_1 &= \sqrt{2a}\left(\cos\frac{\vartheta+\pi}{2} + i\sin\frac{\vartheta+\pi}{2}\right), \\ w_2 &= -\sqrt{2a}\left(\cos\frac{\vartheta+\pi}{2} + i\sin\frac{\vartheta+\pi}{2}\right). \end{aligned} \tag{48}$$

It is evident now that if we take w_1 and go round the circle once so that ϑ changes by the amount 2π, the phase of w_1 changes by π and w_1 becomes equal to w_2. Our approximation does not affect the argument since the interchange of values comes from the factor $(1 - z)$ and not from $(1 + z)$. Going round this particular circle once, the phase of $(1 + z)^{1/2}$ returns to its original value; but the phase of $(1 - z)^{1/2}$ does not. A similar situation exists in the vicinity of $z = -1$. Thus we have shown that (43) cannot represent w everywhere in a region including a branch point. It is really an accurate description of the situation to say that the power series diverges there in "self-defense."

Consider z going round a circle centered at the origin, inside the circle passing through the branch points as shown in Figure 4.3; $(z + 1)$ is represented by the line leading from point (-1) to point z and $(z - 1)$ by the line joining point $(+1)$ to point z. As we go round the circle, the phases of $(z + 1)$ and $(z - 1)$ return to their original values; ph w also returns to its original value; hence w_1, w_2 remain separate. In constructing (43) we have started with separate values for w at $z = 0$; as the z-circle expands the series continue to represent w_1 and w_2 until the circle through the branch points is reached. This is the limit of convergence; for if the series converged in a larger circle, we should be faced with a contradiction: in small regions surrounding each branch point we should have w behaving at once as one multiple-valued function and as two single-valued functions.

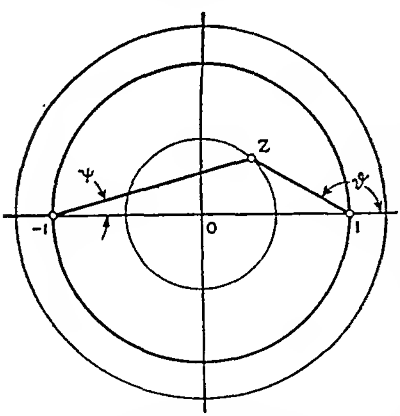

FIG. 4.3. As point z moves once round any circle concentric with the origin, except the unit circle, the two branches of (39) are not interchanged.

If the z-circle is outside the unit circle, w_1 and w_2 are still separate from each other. The phases of $(z + 1)$ and $(z - 1)$ change by the amount 2π, the phase of their product changes by 4π and the phase of w is altered by 2π; but this does not alter w. Outside the unit circle we shall have a pair

of series of negative powers to represent w. To obtain these series we may proceed as follows

$$\begin{aligned} w &= \sqrt{1-z^2} = \sqrt{-z^2(1-z^{-2})} = \pm iz\sqrt{1-z^{-2}} \\ &= \pm iz(1 - \tfrac{1}{2}z^{-2} - \tfrac{1}{8}z^{-4} - \tfrac{1}{16}z^{-6} - \tfrac{5}{128}z^{-8} - \cdots). \end{aligned} \tag{49}$$

The series are convergent outside the unit circle.

The treacherous behavior of multiple-valued functions is now fully exhibited. Either inside or outside the unit circle, w as given by (39) is represented by a pair of separate single-valued functions. All series in question happen to be absolutely convergent on the unit circle itself; and yet in the entire plane, w is not separable into two single-valued functions. If one of series (49) agrees with one of (43) somewhere on the circle, it should not be taken for granted that the agreement extends to the entire circle. Take the upper sign in (43) at $z = i$; then

$$w(i) = 1 + \tfrac{1}{2} - \tfrac{1}{8} + \tfrac{1}{16} - \tfrac{5}{128} + \cdots. \tag{50}$$

This value agrees with that obtained from (49) if we choose the lower sign. At point $z = -i$, however, the upper sign in (43) corresponds to the upper sign in (49). In general, over the upper half-circle the correspondence between (43) and (49) is one way, and over the lower half-circle the opposite way.

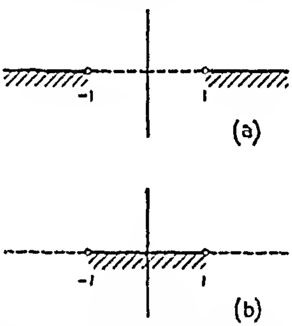

FIG. 4.4. Any multiple-valued function can be represented by a collection of single-valued functions, the *branches* of the multiple-valued function, if we introduce *cuts* which the independent variable is forbidden to cross. These cuts must connect the branch points; otherwise, to a large extent the cuts are arbitrary: (a) and (b) are two possible choices for the function w of equation (39).

The only way to break multiple-valued functions into single-valued functions and thus render them harmless is by means of *cuts* in the complex plane, that is, barriers which are not to be crossed. These cuts should join branch points. A cut could be made over the lower half of the unit circle in Figure 4.3, or along the real axis as in Figure 4.4. In the left part of the figure the cut joins the branch points via the point at infinity (here is where the Neumann sphere is helpful). If the cut is over the lower half of the unit circle, series (43) and (49) can be matched in pairs over the upper half-circle and we have complete separation of the two branches of (39) at the expense of introducing a line of discontinuity over the cut; for there each single-valued function joins continuously not to itself but to the mate.

The problem of multiple-valued functions cannot be dodged and will come up again in Chapters 14 and 15. Serious errors may be made when

insufficient attention is accorded to the treacherous behavior of multiple-valued functions. Such functions can be vanishingly small and extremely large in the vicinity of the same point.

Problem. Study $w = \sqrt{z-1}$. Show that $z = 1$ and $z = \infty$ are the two branch points of w (*Hint:* To establish the behavior of w at infinity, let $z = 1/t$). Obtain the ascending and descending power series in z. Note that there will be a fractional power as a factor in the series of descending powers which renders them multiple-valued so that one of them can represent w everywhere outside the unit circle.

Ans.
$$w = \pm i(1 - \tfrac{1}{2}z - \tfrac{1}{8}z^2 - \tfrac{1}{16}z^3 - \tfrac{5}{128}z^4 - \cdots), \qquad |z| < 1;$$
$$= z^{1/2}(1 - \tfrac{1}{2}z^{-1} - \tfrac{1}{8}z^{-2} - \tfrac{1}{16}z^{-3} - \tfrac{5}{128}z^{-4} - \cdots), \qquad |z| > 1.$$

13. *Analytic continuation*

Consider the geometric series (20); the series and all its derivatives converge inside the unit circle. Select some point $z = z_1$ inside the circle and use it as the center of expansion in powers of $(z - z_1)$. The convergence circle of the new series may be either completely inside the former

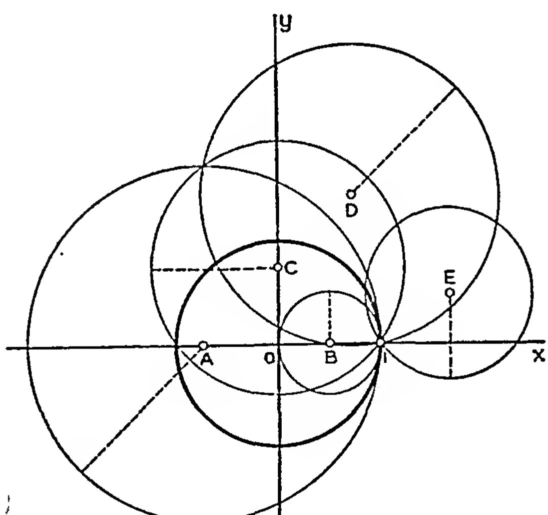

FIG. 4.5. Illustrating the analytic continuation of a function.

circle or partly outside. In the latter case we have succeeded in "continuing" the function outside its previous region of definition. If, in the case of (20), we expand about $A(-3/4)$, Figure 4.5, the new circle of convergence takes us outside the former circle; if we expand about $B(1/2)$, it will not. The reason is: every circle of convergence will pass through the singular point nearest to the center of expansion. Nevertheless, unless the

boundary of the original circle is so dense with singular points that we cannot squeeze out between, we can gradually extend the region of definition of our function.

The example (20) is trivial in the sense that we really know that it represents (22) and we can find directly all the power series we want

$$\frac{1}{1-z} = \frac{1}{(1-z_1)-(z-z_1)} = \frac{1}{1-z_1}\cdot\frac{1}{1-(z-z_1)/(1-z_1)}$$
$$= \frac{1}{1-z_1} + \frac{z-z_1}{(1-z_1)^2} + \frac{(z-z_1)^2}{(1-z_1)^3} + \cdots + \frac{(z-z_1)^n}{(1-z_1)^{n+1}} + \cdots. \tag{51}$$

The radius of convergence of this series is $|1 - z_1|$. In this particular case we do not really want the various expansions since we have a simple analytic expression which is good everywhere; but we should think of those occasions when the power series is the only source of knowledge of the function.

Now, there is no law against our defining a function as follows

$$\begin{aligned} f(z) &= \frac{1}{1-z}, \quad \text{if } |z| < 1, \\ &= z, \quad \text{if } |z| \geq 1. \end{aligned} \tag{52}$$

The analytic continuation of this function from the inside of the unit circle will obviously not yield its values outside the unit circle. This is because the process of analytic continuation involves the assumption of the existence of all derivatives of the function on the unit circle. The derivatives of (52) do not exist on the unit circle and so far as this function is concerned the unit circle is an impenetrable barrier of singular points.

Sometimes it is impossible to tell, by a superficial inspection, whether or not a given function can be extended; the following series

$$w = 1 + z + z^2 + z^4 + z^8 + z^{16} + \cdots + z^{2^n} + \cdots \tag{53}$$

is an illustration. The series looks all right; in fact, it is just like (20) with most terms missing; but all attempts to continue it will fail for the boundary of convergence is packed with singularities.

14. *Asymptotic expansions*

The *sine integral* defined by

$$\mathrm{Si}\, x = \int_0^x \frac{\sin t}{t}\, dt \tag{54}$$

occurs in radiation theory. If the integrand is expanded in a power series and integrated termwise, a series for Si x is obtained which converges for all

values of x. However, for large x convergence is slow; in this region a much more serviceable series can be obtained as follows.

From (54) we have

$$\operatorname{Si} x = \int_0^\infty \frac{\sin t}{t}\,dt - \int_x^\infty \frac{\sin t}{t}\,dt. \tag{55}$$

The second integral is integrated by parts

$$\begin{aligned}\int_x^\infty \frac{\sin t}{t}\,dt &= \int_x^\infty (1/t)d(-\cos t) = -\frac{\cos t}{t}\Big]_x^\infty + \int_x^\infty \cos t\, d(1/t) \\ &= \frac{\cos x}{x} - \int_x^\infty \frac{\cos t}{t^2}\,dt.\end{aligned} \tag{56}$$

Continuing the process, we obtain

$$\begin{aligned}\int_x^\infty \frac{\sin t}{t}\,dt &= \frac{\cos x}{x} + \frac{\sin x}{x^2} - \frac{2!}{x^3}\cos x - \frac{3!}{x^4}\sin x + \cdots \\ &= \frac{\cos x}{x}\sum_{n=0}^{\infty}\frac{(-)^n(2n)!}{x^{2n}} + \frac{\sin x}{x}\sum_{n=0}^{\infty}\frac{(-)^n(2n+1)!}{x^{2n+1}}.\end{aligned} \tag{57}$$

Each of these series diverges for all values of x. In the first series, for instance, the ratio of the nth term to the $(n-1)$th is $-2n(2n-1)/x^2$ and if n is large enough this ratio exceeds unity (in absolute value) no matter how large is x. On the other hand, if we take *a fixed number* of terms, then as x increases the finite series will approximate the given integral better and better. This may be anticipated from (56) where the integrand on the right is smaller than the original integrand (if x is greater than unity); after the next integration by parts, the integral representing the remainder of the series will be still smaller if x is large enough.

We are now ready for a formal definition. An infinite descending power series is said to be *asymptotic* to $f(x)$ for large x and this relationship is expressed symbolically by writing

$$f(x) \sim a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \cdots + \frac{a_n}{x^n} + \cdots \tag{58}$$

if, for a fixed n,

$$\lim x^n \left[f(x) - a_0 - \frac{a_1}{x} - \frac{a_2}{x^2} - \cdots - \frac{a_n}{x^n}\right] = 0 \text{ as } x \to \infty. \tag{59}$$

If an infinite power series is expressed as a finite series with the remainder term $R(x,n)$ depending on the variable x and the number of terms n, then in order that the series be convergent R should approach zero with increasing *n for a fixed value of x*; and, in order that the series be asymptotic, R should approach zero with increasing *x for a fixed value of n*.

The definition of series asymptotic to $f(x)$ is such that all functions $f(x) + F(x)$ have the same asymptotic series if $F(x)$ decreases with increasing x faster than x^{-n}; for then $\lim x^n F(x) = 0$. An example of such $F(x)$ is $\exp(-x)$. Normally, this ambiguity does not cause any real difficulty.

Problems

1. The *error function and the complementary error function* are defined as follows

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt, \qquad \operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\,dt.$$

Expand erf x in a power series and erfc x in an asymptotic series.

Ans.

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} (-)^n \frac{x^{2n+1}}{n!(2n+1)},$$

$$\operatorname{erfc} x \sim \frac{e^{-x^2}}{x\sqrt{\pi}} \left[1 - \frac{1}{2x^2} + \frac{1 \cdot 3}{(2x^2)^2} - \frac{1 \cdot 3 \cdot 5}{(2x^2)^3} + \cdots\right].$$

Note: In Chapter 6 it is shown that

$$\operatorname{erf} x + \operatorname{erfc} x = 1.$$

2. *Fresnel integrals* are

$$C(x) = \int_0^x \cos(\pi t^2/2)\,dt, \qquad S(x) = \int_0^x \sin(\pi t^2/2)\,dt.$$

Obtain convergent power series and asymptotic series, using $C(\infty) = S(\infty) = \frac{1}{2}$.
Note: The answers are given in Chapter 19.

15. *Power series and Fourier series*

As shown by (26), two Fourier series can be obtained from any power series. Substituting the polar form of z in the geometric series (20) and its sum (22), we have

$$\begin{aligned} \frac{1}{(1 - \rho\cos\varphi) - i\rho\sin\varphi} &= 1 + \rho\cos\varphi + \rho^2\cos 2\varphi + \rho^3\cos 3\varphi + \cdots \\ &\quad + i(\rho\sin\varphi + \rho^2\sin 2\varphi + \rho^3\sin 3\varphi + \cdots). \end{aligned} \tag{60}$$

Multiplying both numerator and denominator of the fraction by the conjugate of the denominator and separating the real and imaginary parts, we get

$$\begin{aligned} S &= \frac{1 - \rho\cos\varphi}{1 - 2\rho\cos\varphi + \rho^2} = 1 + \rho\cos\varphi + \rho^2\cos 2\varphi + \rho^3\cos 3\varphi + \cdots, \\ T &= \frac{\rho\sin\varphi}{1 - 2\rho\cos\varphi + \rho^2} = \rho\sin\varphi + \rho^2\sin 2\varphi + \rho^3\sin 3\varphi + \cdots. \end{aligned} \tag{61}$$

Letting

$$k = 2\rho/(1 + \rho^2), \qquad \sqrt{1 - k^2} = (1 - \rho^2)/(1 + \rho^2), \tag{62}$$

and dividing both terms of the fractions in (61) by $(1 + \rho^2)$, we obtain

$$S - \frac{1}{2} = \frac{1 - \rho^2}{2(1 - 2\rho\cos\varphi + \rho^2)} = \frac{\sqrt{1 - k^2}}{2(1 - k\cos\varphi)},$$
$$T = \frac{k\sin\varphi}{2(1 - k\cos\varphi)}. \tag{63}$$

Therefore

$$\frac{1}{1 - k\cos\varphi} = \frac{2}{\sqrt{1 - k^2}}\left(\frac{1}{2} + \rho\cos\varphi + \rho^2\cos 2\varphi + \cdots\right),$$
$$\frac{\sin\varphi}{1 - k\cos\varphi} = \frac{2}{k}(\rho\sin\varphi + \rho^2\sin 2\varphi + \rho^3\sin 3\varphi + \cdots), \tag{64}$$
$$\rho = \frac{1 - \sqrt{1 - k^2}}{k} = \frac{k}{1 + \sqrt{1 - k^2}}.$$

Since ρ in (61) must be smaller than unity, k is also smaller than unity; this should be taken into account in obtaining ρ from (62).

Problems

1. Obtain

$$\frac{1 - 2\rho\cos\varphi + \rho^2\cos 2\varphi}{(1 - 2\rho\cos\varphi + \rho^2)^2} = 1 + 2\rho\cos\varphi + 3\rho^2\cos 2\varphi + \cdots,$$
$$\frac{2\rho\sin\varphi - \rho^2\sin 2\varphi}{(1 - 2\rho\cos\varphi + \rho^2)^2} = 2\rho\sin\varphi + 3\rho^2\sin 2\varphi + 4\rho^3\sin 3\varphi + \cdots.$$

Hint: The series may be obtained from the power series for $(1 - z)^{-2}$.

2. Obtain

$$S = \sqrt{\tfrac{1}{2}(1 - \rho\cos\varphi + \sqrt{1 - 2\rho\cos\varphi + \rho^2})}$$
$$= 1 - \tfrac{1}{2}\rho\cos\varphi - \tfrac{1}{8}\rho^2\cos 2\varphi - \tfrac{1}{16}\rho^3\cos 3\varphi - \tfrac{5}{128}\rho^4\cos 4\varphi - \cdots,$$
$$\frac{\sin\varphi}{S} = \sin\varphi + \tfrac{1}{4}\rho\sin 2\varphi + \tfrac{1}{8}\rho^2\sin 3\varphi + \tfrac{5}{64}\rho^3\sin 4\varphi + \cdots.$$

CHAPTER V

Differentiation

There are many kinds of derivatives: ordinary derivatives of a function, $y = f(x)$, of one independent variable; partial derivatives of a function of several independent variables; total derivatives; directional derivatives; derivatives, such as the divergence and curl, of sets of functions. The essential idea is always the same, although there are variations with the circumstances. The formal process of calculation is always the same; it consists primarily of obtaining one or more ordinary derivatives of one or more functions of one independent variable. For this reason we shall concentrate our attention on definitions.

1. *Pictorial representation of functions*

In Figure 5.1 we have three samples of the most common graphic representation of a real function of one real independent variable, $y = f(x)$.

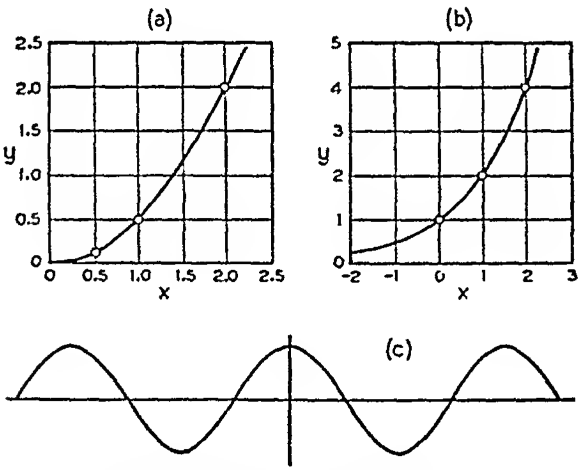

Fig. 5.1. Samples of pictorial representation of functions: (a) $y = \frac{1}{2}x^2$, (b) $y = 2^x$, (c) $y = \cos x$.

Figure 5.2 illustrates another method of representation which is particularly suitable for such rapidly varying functions as $y = 10^x$.

Functions, $z = f(x,y)$, of two independent variables pose a problem. It is easy enough to think of x, y, z as cartesian coordinates of a point in space;

then $z = f(x,y)$ is represented by a surface. Models can be made and photographs taken; but there are obvious disadvantages to this. Another method is preferable. Suppose that z is considered as the height above the sea level of a point on a plane earth; then we can draw *contour* or *level lines*

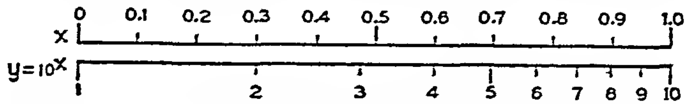

FIG. 5.2. The "double scale" method of pictorial representation of $y = 10^x, 0 \leq x \leq 1$.

connecting points of constant z. If the successive increments in z are constant, such a " survey map " of the function gives a clear indication of the mountains, valleys, ravines, and " saddle points " or " passes " of the function.

For example, in Figure 5.3 the level lines are the " equipotential lines " of two equally charged filaments, passing through A and B and perpendicu-

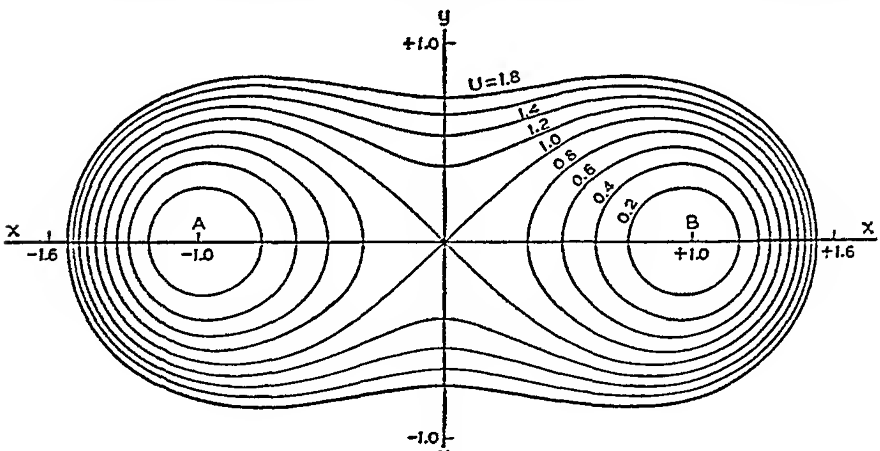

FIG. 5.3. The "survey map" of the function $U = [(1 + x)^2 + y^2] [(1 - x)^2 + y^2]$. The level lines of this function coincide with the equipotential lines surrounding two parallel electrically charged filaments, passing through A and B at right angles to the plane of the paper, in the special case of equal and opposite electrification. The potential is proportional to log U.

lar to the plane of the paper. The saddle point is at the origin; it is a point of minimum potential along the x-axis and maximum potential along the y-axis.

This type of graphic representation is not feasible when the number of independent variables exceeds two; still, it is convenient to retain the language of the representation. Thus it is usual to refer to level surfaces of a function of three independent variables. Frequently these variables are actually the coordinates of a point in space; in any case, they can be

regarded as such. In special instances, level surfaces are given distinctive names; thus, there are equipotential, isothermal, isobaric, etc., surfaces.

2. *Average derivative*

The *average rate of change* of a function $y = f(x)$ in the interval from $x = x_1$ to $x = x_2$, or the *average derivative* of y, is the ratio of the increment $\Delta y = f(x_2) - f(x_1)$ of the dependent variable to the increment $\Delta x = x_2 - x_1$ of the independent variable. For example, the average derivative of $y = x^2$ in the interval (x_1,x_2) is

$$\frac{\Delta y}{\Delta x} = \frac{x_2^2 - x_1^2}{x_2 - x_1} = x_1 + x_2; \qquad (1)$$

in the interval $(x,x + \Delta x)$ it is

$$\frac{\Delta y}{\Delta x} = \frac{(x + \Delta x)^2 - x^2}{\Delta x} = 2x + \Delta x. \qquad (2)$$

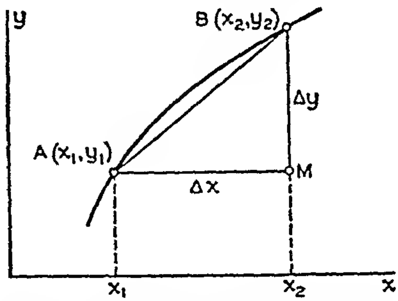

FIG. 5.4. The *increments* of the independent and dependent variables are the differences between these variables at two points.

On a graph of $y = f(x)$ the average derivative is represented by the slope of the chord AB joining the ends of the interval, provided of course that the vertical and horizontal scales are the same, Figure 5.4.

Problems

1. Show that $x_1^2 + x_1x_2 + x_2^2$ is the average derivative of $y = x^3$ in the interval (x_1,x_2).

2. Show that $\cos \frac{1}{2} (x_1 + x_2) \dfrac{\sin \frac{1}{2}(x_1 - x_2)}{\frac{1}{2}(x_1 - x_2)}$ is the average derivative of $\sin x$ in the interval (x_1,x_2).

3. Show that $- \sin (x + \frac{1}{2} \Delta x) \dfrac{\sin \frac{1}{2}\Delta x}{\frac{1}{2}\Delta x}$ is the average derivative of $\cos x$ in the interval $(x,x + \Delta x)$.

4. Compute the average derivative of $y = J_0(x)$ in the interval from $x = 1.2$ to $x = 1.3$. (*Note:* $J_0(x)$ is the symbol for the Bessel function of the first kind of order zero.) *Ans.* -0.510.

5. Show that the average derivative of $\sqrt{x}$ in the interval (x_1,x_2) is

$$1/(\sqrt{x_1} + \sqrt{x_2}).$$

3. *Derivative*

The *derivative* Dy of $y = f(x)$ is the limit of the average derivative $\Delta y/\Delta x$ as Δx approaches zero; thus

$$Dy = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \qquad (3)$$

For instance, the derivative of $y = x^2$ is $Dy = 2x$ and the derivative of $y = x^3$ is $Dy = 3x^2$. Similarly $D(\sin x) = \cos x$ and $D(\cos x) = -\sin x$.

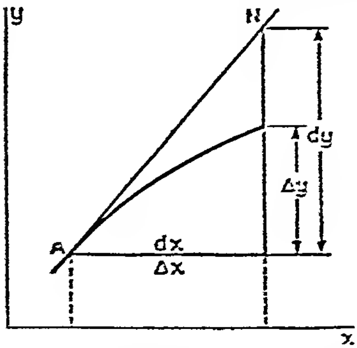

FIG. 5.5. The *differential* dx of the independent variable equals the increment Δx; the differential dy of the dependent variable is the increment along the tangent and normally *is not* equal to the increment Δy.

Geometrically Dy is represented by the slope of the tangent AN to the curve $y = f(x)$, Figure 5.5.

4. *Differential*

By definition the limit of $\Delta y/\Delta x$ is Dy; hence

$$\frac{\Delta y}{\Delta x} = Dy + \epsilon, \tag{4}$$

where ϵ approaches zero together with Δx. Equation (4) may be written

$$\Delta y = (Dy)\Delta x + \epsilon\Delta x. \tag{5}$$

Except when $Dy = 0$ the second term on the right approaches zero faster than the first and the principal part

$$dy = (Dy)\Delta x = (Dy)\,dx \tag{6}$$

is called the *differential of y*. There is no difference between the increment and the differential of the independent variable; if $y = x$, then, by definition, $dy = \Delta x$ and, therefore, $dx = \Delta x$.

Neither increments nor differentials are required to be infinitesimal; in applied mathematics, however, it is frequently assumed that they are.

The most common notation for the derivative, dy/dx, follows from (6); thus

$$Dy = \frac{dy}{dx}. \tag{7}$$

The symbol Dy was introduced by Cauchy and dy/dx by Leibnitz.

Problems

1. Show that if $y^2 = x^3$, then $Dy = 3x^2/2y = (3/2)x^{1/2}$.
2. Show that if $\sin y = x^2$, then $Dy = 2x/\cos y$.
3. Show that if $y(x) = u(x)v(x)$, then $Dy = u\,Dv + v\,Du$.
4. Show that if $y(x) = 1/u(x)$, then $Dy = -(Du)/u^2$.
5. Show that if $y(x) = u(x)/v(x)$, then $Dy = (v\,Du - u\,Dv)/v^2$.
6. Show that if $F(y) = f(x)$, then $Dy = f'(x)/F'(y)$ where the primes indicate the derivatives with respect to the indicated arguments.
7. Obtain the derivatives of $\sec x$, $\tan x$, $\cot x$.

Ans. $\tan x \sec x$, $\sec^2 x$, $-\csc^2 x$.

8. Obtain the derivatives of $\sin x^2$, $\cos\sqrt{x}$, $\sin(\tan x^3)$.
Ans. $2x\cos x^2$, $-\frac{1}{2}x^{-1/2}\sin x^{1/2}$, $3x^2\sec^2(x^3)\cos(\tan x^3)$.

5. *Relative derivative*

The *relative increment* of $y = f(x)$ is the increment Δy divided by y itself. The limit of the ratio of the relative increment of y to the increment of x is the *relative rate of change* or the *relative derivative* of y with respect to x. To reiterate:

$$\text{Relative derivative} = \lim_{\Delta x \to 0} \frac{\Delta y}{y\Delta x} = \frac{1}{y}\frac{dy}{dx}. \tag{8}$$

For example, the relative derivative of $y = x^2$ is $2/x$· the relative derivative of $\sin x$ is $\cot x$.

The relative derivative of y is the ordinary derivative of the natural logarithm of y and for this reason it is often called the *logarithmic derivative.* There is much to be said, however, in favor of definition (8), either in pure or applied mathematics. For instance, compare the following two statements of the physical law of radioactive disintegration: (1) the relative rate of disintegration of a radioactive mass is constant and (2) the logarithmic derivative of a radioactive mass is constant. The former statement is more directly intelligible than the latter.

Problems

1. Find the relative derivatives of $\cos x$, $\tan x$, $\sec x$.
Ans. $-\tan x$, $2/\sin 2x$, $\tan x$.
2. Find the relative derivatives of $\sin x^2$, $\cos x^2$, $\sec x^2$.
Ans. $2x\cot x^2$, $-2x\tan x^2$, $2x\tan x^2$.
3. Show that if $u(x) = k/v(x)$, where k is a constant, then the relative derivative of u is equal to the negative of the relative derivative of v.

6. *Partial derivatives and differentials*

Derivatives of functions of several independent variables depend on the conditions imposed on the increments of these variables. If we assume that all increments but one are equal to zero, we have a *partial derivative;* if we assume a linear relationship between the increments, we have a directional derivative (see Section 9).

For example, the partial derivative of $z = f(x,y)$ with respect to x is defined as the derivative of z on the assumption that y is constant; thus

$$D_x z = \frac{\partial z}{\partial x} = \lim_{\Delta x \to 0} \frac{\Delta_x z}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x,y)}{\Delta x}. \tag{9}$$

Similarly. the partial derivative with respect to y is

$$D_y z = \frac{\partial z}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta_y z}{\Delta y} = \lim \frac{f(x,y + \Delta y) - f(x,y)}{\Delta y}. \tag{10}$$

Both notations for partial derivatives are in use but the second is the more common. It should be noted, however, that in the above equations ∂z has no meaning by itself and $\partial z/\partial x$, $\partial z/\partial y$ are not to be regarded as fractions. For this reason $D_x z$ and $D_y z$ might be preferable; but customs die hard, if ever.

From the above equations and the definition of the limit we have the following expressions for the *partial increments*

$$\Delta_x z = \frac{\partial z}{\partial x}\Delta x + \epsilon_1 \Delta x, \qquad \Delta_y z = \frac{\partial z}{\partial y}\Delta_y + \epsilon_2 \Delta y, \tag{11}$$

where ϵ_1 and ϵ_2 approach zero respectively with Δx and Δy. The principal parts of the increments are called *partial differentials*

$$d_x z = \frac{\partial z}{\partial x}\Delta x = \frac{\partial z}{\partial x} dx, \qquad d_y z = \frac{\partial z}{\partial y}\Delta y = \frac{\partial z}{\partial y} dy. \tag{12}$$

Extension of these ideas to functions of more than two independent variables is straightforward.

Problems

1. The expressions for the cartesian coordinates x, y, z of a point P in terms of spherical coordinates are

$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r\cos\theta.$$

Find the partial derivatives of the cartesian coordinates regarded as functions of the spherical coordinates.

Ans. $\dfrac{\partial x}{\partial r} = \sin\theta\cos\varphi, \qquad \dfrac{\partial x}{\partial \theta} = r\cos\theta\cos\varphi, \qquad \dfrac{\partial z}{\partial r} = \cos\theta,$ etc.

2. Find the partial derivatives of the spherical coordinates regarded as functions of the cartesian coordinates.

Ans. $\dfrac{\partial r}{\partial x} = \dfrac{x}{r} = \sin\theta\cos\varphi, \qquad \dfrac{\partial r}{\partial y} = \dfrac{y}{r} = \sin\theta\sin\varphi, \qquad \dfrac{\partial r}{\partial z} = \dfrac{z}{r} = \cos\theta,$

$$\frac{\partial \theta}{\partial x} = \frac{xz}{(x^2 + y^2 + z^2)\sqrt{x^2 + y^2}} = \frac{\cos\theta\cos\varphi}{r}, \quad \text{etc.}$$

Hint: Express r, θ, φ in terms of x, y, z before differentiating.

3. Show that if r is regarded as a function of z, θ, φ, then $\partial r/\partial z = \sec\theta$. Compare this result with the corresponding partial derivative in Problem 2 and note that it is

necessary to choose a complete set of independent variables before we can assign a definite meaning to partial derivatives. The cases in Problems 1 and 2 are the important cases; the one in the present problem is considered only to show the importance of the above remark.

4. Show that if r, x, y are regarded as functions of z, θ, φ, then: $\partial x/\partial z = \tan\theta\cos\varphi$, $\partial x/\partial\theta = z\sec^2\theta\cos\varphi$.

5. Show that if x, θ, φ are regarded as functions of r, y, z, then: $\partial x/\partial z = -z/x$, $\partial\theta/\partial z = -1/r\sin\theta = -(r^2 - z^2)^{-1/2}$.

7. *Total differential*

The *total* or *complete differential* is the sum of all partial differentials; thus in the case of two variables we have

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy. \tag{13}$$

This expression makes it especially clear that $\partial z/\partial x$ and $\partial z/\partial y$ should not be regarded as fractions and that ∂x, ∂y are not dx, dy in disguise.

The *complete increment* of z

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x,y) \tag{14}$$

is not the sum of the partial increments. For example if $z = xy$, then

$$\Delta z = (x + \Delta x)(y + \Delta y) - xy = y\Delta x + x\Delta y + \Delta x\Delta y;$$

but the partial increments are

$$\Delta_x z = y\Delta x \quad \text{and} \quad \Delta_y z = x\Delta y.$$

However, the difference between the complete increment and the sum of all partial increments is an infinitesimal of higher order.

Whenever the total differential is obtained directly from the given function, the coefficients can be identified as partial derivatives. Thus, if $z = x^2y^3$,

$$\begin{aligned}\Delta z = (x + \Delta x)^2(y + \Delta y)^3 - x^2y^3 &= 2xy^3\Delta x + 3x^2y^2\Delta y + 6xy^2\Delta x\Delta y\\ &+ y^3\Delta x^2 + 3x^2y\Delta y^2 + x^2\Delta y^3 + 2x\Delta x\Delta y^3 + \Delta x^2\Delta y^3\\ &+ 3y^2\Delta x^2\Delta y + 6xy\Delta x\Delta y^2 + 3y\Delta x^2\Delta y^2\end{aligned} \tag{15}$$

and

$$dz = 2xy^3\Delta x + 3x^2y^2\Delta y = 2xy^3\,dx + 3x^2y^2\,dy.$$

Therefore,

$$D_x z = \frac{\partial z}{\partial x} = 2xy^3, \qquad D_y z = \frac{\partial z}{\partial y} = 3x^2y^2.$$

Obviously it is not necessary to calculate the increment Δz in full in order to obtain the total differential; in equation (15) the last nine terms could have been lumped together as " higher order terms."

In actual practice there is no difference between the calculation of partial and ordinary derivatives, for in obtaining the former we treat all variables but one as constants and in effect seek ordinary derivatives.

Not every differential expression of the form

$$dV = M(x,y)\,dx + N(x,y)\,dy \tag{16}$$

is the total differential of some function; for instance, $dV = x\,dx + x\,dy$ is not. When the differential expression is the differential of some function, it is often called the *exact* differential. In electrostatics the differential of the electromotive force is an exact differential; but in electrodynamics it is not. Another field in which "inexact" differentials are of common occurrence is thermodynamics.*

If (16) is to be the exact differential of some function $z = f(x,y)$ we must have

$$M(x,y) = \frac{\partial z}{\partial x}, \quad N(x,y) = \frac{\partial z}{\partial y}. \tag{17}$$

These relations are obtained if we compare (16) with (13) and note that dx and dy are completely arbitrary. Differentiating the first of the above equations with respect to y and the second with respect to x, we have the *necessary* condition

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} \tag{18}$$

that (16) be an exact differential.† In the next chapter we shall show that this is also a *sufficient* condition; that is, if (18) is satisfied, a function $z = f(x,y)$ exists for which $dz = M dx + N dy$.

8. *Total derivative*

Suppose that we have a function $V(x,y,z,t)$ in which x, y, z are sometimes independent variables and sometimes functions of t; then there will be occasions for the use of two kinds of derivatives with respect to t: the partial derivative $\partial V/\partial t$ and the *total derivative* dV/dt. As indicated by the

* Leigh Page, *Introduction to Theoretical Physics*, D. Van Nostrand Company, Inc., New York, Second Edition, 1935, Chapter 7. Samuel Glasstone, *Thermodynamics for Chemists*, D. Van Nostrand Company, Inc., New York, 1947. Hugh S. Taylor and Samuel Glasstone, *A Treatise on Physical Chemistry*, Vol. I, Atomistics and Thermodynamics. Samuel Glasstone, *Textbook of Physical Chemistry*, D. Van Nostrand Company, Inc., New York, Second Edition, 1946.

† In this proof it is assumed that the second derivative of z, once with respect to x and once with respect to y, is independent of the order of differentiation. This is indeed the case if $\partial M/\partial y$ and $\partial N/\partial x$ are continuous functions of both independent variables.

notation the total derivative is the ratio of the total differential of the function to the differential of the independent variable; that is,

$$\frac{dV}{dt} = \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt} + \frac{\partial V}{\partial t}. \tag{19}$$

This ratio is well-defined when x, y, z are given functions of t.

If V is the temperature at point (x,y,z) at time t, then $\partial V/\partial t$ is the rate of temperature change as it appears to a fixed observer while dV/dt is the rate as it appears to a moving observer whose position is given by

$$x = x(t), \qquad y = y(t), \qquad z = z(t). \tag{20}$$

Since dx/dt, dy/dt, dz/dt are the components of the velocity of the observer we have

$$\frac{dV}{dt} = v_x\frac{\partial V}{\partial x} + v_y\frac{\partial V}{\partial y} + v_z\frac{\partial V}{\partial z} + \frac{\partial V}{\partial t}. \tag{21}$$

The symbols for the total and ordinary derivatives are the same because as soon as we substitute the given functions from (20), $V(x,y,z,t)$ becomes a function of a single variable t.

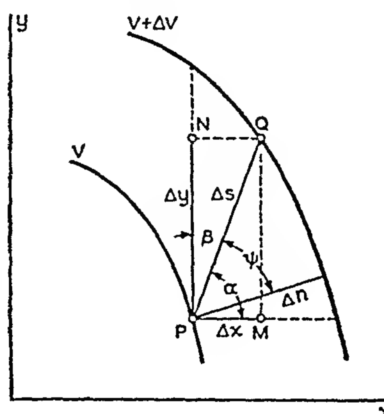

Fig. 5.6. The *directional derivative* is the limit of a quotient in which the numerator is the difference of the values of the function on two level lines (or surfaces) and the denominator is the distance Δs between the level lines in the given direction. The most important preferred directions are along the normal to the level line passing through the given point and along the coordinate lines.

9. *Directional derivatives*

Let $V = f(x,y)$ be represented by level lines in the xy-plane and consider the lines corresponding to V and $V = \Delta V$, Figure 5.6. On these lines choose a pair of points, P and Q. The average rate of change along PQ is $\Delta V/\Delta s$, where Δs is the length PQ. The *derivative at P in the direction PQ* is the limit of the average derivative along PQ as Δs approaches zero

$$\frac{\partial V}{\partial s} = \lim \frac{\Delta V}{\Delta s}, \quad \Delta s \to 0. \tag{22}$$

If $V = x^2 + y^2$, the level lines form a family of circles concentric with the origin. The increment of V in passing from $P(x,y)$ to $Q(x + \Delta x, y + \Delta y)$ is

$$\Delta V = 2x\Delta x + 2y\Delta y + \Delta x^2 + \Delta y^2,$$

and since $PQ = \Delta s = \sqrt{\Delta x^2 + \Delta y^2}$ the average derivative is

$$\frac{\Delta V}{\Delta s} = 2x\frac{\Delta x}{\sqrt{\Delta x^2 + \Delta y^2}} + 2y\frac{\Delta y}{\sqrt{\Delta x^2 + \Delta y^2}} + \frac{\Delta x^2 + \Delta y^2}{\sqrt{\Delta x^2 + \Delta y^2}}. \tag{23}$$

If k is the slope of PQ, then $\Delta y = k\Delta x$. Substituting this in (23) and letting Δx approach zero while keeping k constant, we have

$$\frac{\partial V}{\partial s} = \frac{2x + 2ky}{\sqrt{1 + k^2}}. \tag{24}$$

More generally

$$\Delta V = \frac{\partial V}{\partial x}\Delta x + \frac{\partial V}{\partial y}\Delta y + \text{infinitesimals of higher order.} \tag{25}$$

If α and β are the angles made by PQ with the x and y-axes, then

$$\frac{\Delta x}{\Delta s} = \cos\alpha, \qquad \frac{\Delta y}{\Delta s} = \cos\beta. \tag{26}$$

Hence the directional derivative is

$$\frac{\partial V}{\partial s} = \frac{\partial V}{\partial x}\cos\alpha + \frac{\partial V}{\partial y}\cos\beta. \tag{27}$$

In the present case $\cos\beta = \sin\alpha$; but the symmetrical form (27) is the one we obtain when the number of independent variables is greater than two Thus, if $V = f(x,y,z)$, then

$$\frac{\partial V}{\partial s} = \frac{\partial V}{\partial x}\cos\alpha + \frac{\partial V}{\partial y}\cos\beta + \frac{\partial V}{\partial z}\cos\gamma, \tag{28}$$

where α, β, γ are the angles made with the coordinate axes by a typical direction in the 3-space.

Partial derivatives are seen to be the derivatives in the directions of the coordinate axes.

10. *Gradient*

From the geometric representation it is evident that the magnitude of the derivative in the direction of the normal to level lines (or level surfaces in the 3-space) is maximum, for in that direction Δs becomes eventually smaller than the corresponding Δs in any other direction. This maximum derivative is called the *gradient* of V and in Vector Analysis it is denoted by "grad V." The notation is supposed to include both the magnitude and direction. The magnitude alone is denoted by $\partial V/\partial n$, "n" for normal.

If ψ is the angle between PQ and the normal, then

$$\Delta n = (\Delta s) \cos\psi + \text{higher order infinitesimals.}$$

Therefore,

$$\frac{\partial V}{\partial n} = \frac{\partial V}{\partial s}\sec\psi \quad \text{or} \quad \frac{\partial V}{\partial s} = \frac{\partial V}{\partial n}\cos\psi. \tag{29}$$

Thus we have another expression for the directional derivative as well as a proof that the magnitude of the normal derivative is really maximum. There are, of course, two normal derivatives corresponding to the two directions of the normal; if we decide to fix one particular direction as "positive," the negative normal will be specified by $\psi = \pi$.

If A, B, C are the angles between the coordinate axes and the normal, then from (29) we obtain

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial n}\cos A, \qquad \frac{\partial V}{\partial y} = \frac{\partial V}{\partial n}\cos B, \qquad \frac{\partial V}{\partial z} = \frac{\partial V}{\partial n}\cos C. \tag{30}$$

Squaring, adding, and taking the square root of the result, we have

$$\frac{\partial V}{\partial n} = \sqrt{\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 + \left(\frac{\partial V}{\partial z}\right)^2}. \tag{31}$$

For any number of independent variables, the magnitude of the normal derivative is the square root of the sum of the squares of all partial derivatives.

Problems

1. Find the derivative of $V = 3xy + y^2$ in the direction parallel to $y = kx$.
Ans. $[3y + k(3x + 2y)]/\sqrt{1 + k^2}$.

2. Find the maximum directional derivative of the function given in Problem 1.
Ans. $(9x^2 + 12xy + 13y^2)^{1/2}$.

3. Find the direction in which $\partial V/\partial s$, where V is given in Problem 1, is maximum.
Ans. Parallel to $Y = \left(\frac{x}{y} + \frac{2}{3}\right)X$.

4. What are the answers to the preceding problems if $V = x^3 + 2xy - y^2$.
Ans. $[3x^2 + 2y + 2k(x - y)]/\sqrt{1 + k^2}$; $(9x^4 + 12x^2y + 4x^2 - 8xy + 8y^2)^{1/2}$; parallel to $Y = \frac{2(x - y)}{3x^2 + 2y}X$.

5. Find the derivative of $V = x^2 + yz$ in the direction of the line joining the origin with point (2,1,3), the normal derivative, and the "direction cosines" of the normal.

6. Show that (28), (29) and (30) imply that the angle ψ between any two directions given by direction cosines $(\cos A, \cos B, \cos C)$ and $(\cos\alpha, \cos\beta, \cos\gamma)$ is given by

$$\cos\psi = \cos A\cos\alpha + \cos B\cos\beta + \cos C\cos\gamma.$$

11. *Derivatives of functions of a complex variable*

First let us consider an example of a function w of a complex variable z in which w is obtained by performing an arithmetic operation upon z

$$w = z^2. \tag{32}$$

In this case the procedure for calculating the derivative is exactly the same as in the case of corresponding functions of a real variable

$$\begin{aligned} \Delta w &= (z + \Delta z)^2 - z^2 = 2z\Delta z + \Delta z^2, \\ \Delta w/\Delta z &= 2z + \Delta z, \quad dw/dz = \lim\,(\Delta w/\Delta z) = 2z. \end{aligned} \tag{33}$$

The derivative is clearly a function of z and is entirely independent of Δz. There is no doubt about it if w is expressed in the form (32); but suppose we were given explicitly the real and imaginary parts

$$w = (x^2 - y^2) + 2ixy. \tag{34}$$

It is no longer clear that the derivative will be independent of the manner in which $\Delta z = \Delta x + i\Delta y$ approaches zero. In fact, the opposite will usually be true. For instance, consider the function

$$w = (x^2 + y^2) + 2ixy,$$

$$\Delta w = 2x\Delta x + \Delta x^2 + 2y\Delta y + \Delta y^2 + 2iy\Delta x + 2ix\Delta y + 2i\Delta x\Delta y, \tag{35}$$

$$\frac{dw}{dz} = \lim \frac{\Delta w}{\Delta z} = \lim \frac{\Delta w}{\Delta x + i\Delta y} = \frac{2x + 2y(dy/dx) + 2iy + 2ix(dy/dx)}{1 + i(dy/dx)},$$

provided Δz approaches zero in such a way that $\Delta y/\Delta x$ approaches a limit, that is, provided the point z is approached along some curve which has a tangent at point z.

Some functions of a complex variable have derivatives depending only on the independent variable and not on the direction of approach to the point in question. Such functions are called *monogenic functions.* Those functions which are given by the same set of arithmetic operations on each value of the variable z *as a whole* are clearly monogenic functions; and it can be shown that all monogenic functions can be represented in this way. Monogenic functions are also called *analytic functions* for, if they are given over even a small portion of a curve in the z-plane, they can be determined throughout the entire plane with the possible exception of some " singular " points. Other functions of a complex variable are called *polygenic functions.* It is monogenic functions, however, which play such an important role in applied mathematics.

We shall now derive the *Cauchy-Riemann* conditions, satisfied by monogenic functions alone. Generally we should say that $w = F(z)$ is a

function of z if the real and imaginary parts of w are functions of the real and imaginary parts of z

$$u = u(x,y), \qquad v = v(x,y). \tag{36}$$

The differentials are

$$\begin{aligned} dz &= dx + i dy, \\ dw &= \left(\frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy\right) + i\left(\frac{\partial v}{\partial x} dx + \frac{\partial v}{\partial y} dy\right); \end{aligned} \tag{37}$$

and their ratio is

$$\frac{dw}{dz} = \frac{\left(\dfrac{\partial u}{\partial x} + i\dfrac{\partial v}{\partial x}\right) + \left(\dfrac{\partial u}{\partial y} + i\dfrac{\partial v}{\partial y}\right)\dfrac{dy}{dx}}{1 + i\dfrac{dy}{dx}} \tag{38}$$

Thus the derivative depends, in general, on the direction $k = dy/dx$ of approach to point z.

This derivative can be represented by points in another complex plane, the *derivative plane*. In Section 1.11 it is shown that the locus of these points, as k varies from $-\infty$ to ∞, is a circle (note that k is a *real* parameter). The center of the circle is at

$$C = \frac{1}{2}\left(\frac{\partial v}{\partial y} + \frac{\partial u}{\partial x}\right) + \frac{1}{2} i\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right); \tag{39}$$

the square of the radius is

$$r^2 = \frac{1}{4}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)^2 + \frac{1}{4}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2. \tag{40}$$

The function is monogenic if the radius of the derivative circle vanishes. This happens if

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{41}$$

These are the Cauchy-Riemann equations; if they are fulfilled,

$$\frac{dw}{dz} = C = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}. \tag{42}$$

Both u and v satisfy the following second order partial differential equation (Laplace's equation)

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \qquad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0. \tag{43}$$

If u is known, v can be determined from the Cauchy-Riemann equations and vice versa. For example, if $u = x^2 - y^2$,

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} = 2x, \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} = 2y. \tag{44}$$

From the first equation, we have

$$v = 2xy + \text{a function of } x \text{ only;}$$

from the second

$$v = 2xy + \text{a function of } y \text{ only.} \tag{45}$$

Therefore,

$$v = 2xy + \text{a constant.} \tag{46}$$

On the other hand, if $u = x^2 + y^2$, the same procedure leads to the conclusion that the Cauchy-Riemann equations have no solution and u is not the real part of any monogenic function. That is, the real part of a monogenic function determines the imaginary part (except for a constant) and vice versa; but the real and imaginary parts of polygenic functions are independent of each other.

Geometrically a function of a complex variable defines a correspondence between the points in two planes. One of these planes may be regarded as a map of the other. In Chapter 14 it is shown that monogenic functions preserve shapes of infinitesimal figures and thus may be used for *conformal mapping*.

Problems

1. Show by direct calculation that $w = x - iy$ is a polygenic function of $z = x + iy$, that the radius of its derivative circle is unity, and that the phase of the derivative is $-2 \tan^{-1} (dy/dx)$.

2. Show by direct calculation that $w = (x^3 - 3xy^2) + i(3x^2y - y^3)$ is a monogenic function of $z = x + iy$ and that the derivative is $3(x^2 - y^2) + 6ixy$.

3. Using the Cauchy-Riemann equations obtain the imaginary part of w in Problem 2 from the real part and vice versa.

4. Show that (39) is equivalent to

$$C = \frac{1}{2}\left(\frac{\partial w}{\partial x} - i\frac{\partial w}{\partial y}\right).$$

5. Let $\dfrac{dy}{dx} = \tan \theta$ in (38). Show that

$$\frac{dw}{dz} = \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right) + \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} + i\frac{\partial u}{\partial y} - \frac{\partial v}{\partial y}\right)e^{-2i\theta}$$

$$= \frac{1}{2}\left(\frac{\partial w}{\partial x} - i\frac{\partial w}{\partial y}\right) + \frac{1}{2}\left(\frac{\partial w}{\partial x} + i\frac{\partial w}{\partial y}\right)e^{-2i\theta}.$$

This represents a vector, $\frac{1}{2}\left(\frac{\partial w}{\partial x}+i\frac{\partial w}{\partial y}\right)$, *rotating clockwise about the point* $\frac{1}{2}\left(\frac{\partial w}{\partial x}-i\frac{\partial w}{\partial y}\right)$ with the angular velocity 2θ.

6. Obtain equations (39), (40), (41), (42) from the results in the preceding problem.

12. *Divergence and curl*

Two other kinds of derivatives, the divergence and curl of a vector, are explained in Chapter 7. These derivatives are associated not with single functions but with sets of functions or " components " of a vector.

CHAPTER VI

INTEGRATION

1. *Integration*

The term *integration* possesses a double meaning — at least when the concept is first introduced in elementary calculus. After learning how to find the derivatives of various functions, we begin to wonder about reversing the process and ask ourselves the question: What is the function whose derivative is given? The most obvious and tempting thing to do would be to prepare a table of derivatives and interchange the columns so that we could read off directly the " inverse derivatives " or " antiderivatives " of given functions. From this point of view a natural symbol for the antiderivative is D^{-1}; thus, if $f(x)$ is the derivative of $y(x)$, we should say that $y(x)$ is an antiderivative of $f(x)$. Symbolically:

$$\text{if } Dy(x) = f(x), \qquad \text{then} \qquad y(x) = D^{-1}f(x). \tag{1}$$

For example,

$$D(x^n) = nx^{n-1} \qquad \therefore \qquad x^n = D^{-1}(nx^{n-1});$$

or, if $n \neq$,

$$D^{-1}(x^{n-1}) = \frac{1}{n}x^n.$$

On the ght-hand side an arbitrary constant may be added since the derivati of a constant is zero.

This int of view is simple but not entirely satisfactory. Thus the above exaple gives no answer to the question: What is the antiderivative of $1/x$? If sne clever person suggests that the answer is $\log x$, we might go on to ask hi for an antiderivative of $(\log x)^{1/2}$.

Going back to e original question, let us denote by $y(x)$ the unknown function whose derivtive $f(x)$ is given. Our problem is then to solve the following *differential eqtion*

$$\frac{dy}{dx} = f(x) \qquad \text{or} \qquad dy = f(x)\,dx. \tag{2}$$

This equation gives very deite instructions for computing the unknown function to any desired degr of accuracy in terms of its value at some initial point. Equation (2) is an abbreviation of

$$\Delta y = f(x)\Delta x + \epsilon\Delta x, \tag{3}$$

where ϵ approaches zero together with Δx; hence if x differs but little from some particular value a, and if we assume that $y(a)$ is known, then

$$y(x) - y(a) \simeq f(a)(x - a), \quad \text{or} \quad y(x) \simeq y(a) + f(a)(x - a). \tag{4}$$

The accuracy of this equation improves as $x - a$ becomes smaller.

Let h be a small quantity. In the interval $a \leq x \leq a + h$ we approximate $y(x)$ as in (4). Replacing a by $a + h$, we find that in the interval $a + h \leq x \leq a + 2h$,

$$y(x) \simeq y(a + h) + f(a + h)(x - a - h),$$

where $y(a + h)$ is found from (4). Repeating the substitution again and again, we obtain

$$y(x) \simeq y(a + nh) + f(a + nh)(x - a - nh), \tag{5}$$

when

$$a + nh \leq x \leq a + (n + 1)h.$$

In each interval $y(x)$ is approximated by a straight line; but the slope of the straight line varies from one interval to the next, Figure 6.1. As the interval h approaches zero, the broken line approaches the limit curve which represents $y(x)$ exactly.

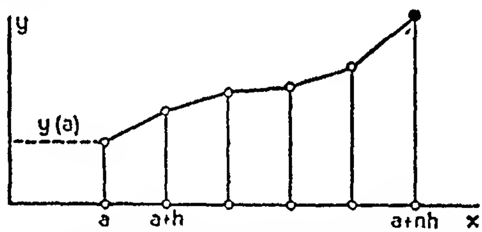

FIG. 6.1. Illustrating a step-by-step method of approximate solution of the differential equation (2) and of evaluation of the definite integral (8).

If $a + nh \leq x \leq a + (n + 1)h$, we have successively

$$\begin{aligned}
y(a + h) &\simeq y(a) + f(a)h, \\
y(a + 2h) &\simeq y(a + h) + f(a + h)h, \\
y(a + 3h) &\simeq y(a + 2h) + f(a + 2h)h, \\
&\cdots\cdots\cdots\cdots \\
y(a + nh) &\simeq y(a + \overline{n - 1}h) + f(a + \overline{n - 1}h)h, \\
y(x) &\simeq y(a + nh) + f(a + nh)(x - a - nh).
\end{aligned} \tag{6}$$

Adding, we obtain

$$\begin{aligned}
y(x) \simeq y(a) + [f(a) + f(a + h) + f(a + 2h) + \cdots \\
+ f(a + \overline{n - 1}h)]h + f(a + nh)(x - a - nh).
\end{aligned} \tag{7}$$

This expression brings out the fact that, in so far as (2) is concerned, the *initial value* $y(a)$ is arbitrary and the problem of solving the equation consists in finding the sum, or rather its limit as h approaches zero. This limit is called the *integral of* $f(x)$ *from* $x = a$ *to* $x = x$; thus

$$\int_a^x f(x)\,dx = \lim \sum_{m=0}^{n-1} f(a + mh)h \tag{8}$$

where $\Delta x \to 0$ and $\lim nh = x - a$. Hence in the limit (7) becomes

$$y(x) = y(a) + \int_a^x f(x)\,dx. \tag{9}$$

It should be noted that the integral (8) is that particular solution of (2) which vanishes at $x = a$.

Equation (9) is a solution of the differential equation (2) in the sense that it implies a definite process for calculating the unknown function $y(x)$. If, however, $y(x)$ is found by some process other than that implied by (8) — from a table of derivatives for instance — then the same equation yields the value of the integral

$$\int_a^x f(x)\,dx = y(x) - y(a). \tag{10}$$

If $x = b$, then

$$\int_a^b f(x)\,dx = y(b) - y(a). \tag{11}$$

The equation shows that the variable x under the integral sign is a "dummy variable" and can be denoted by an arbitrary letter; thus

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(u)\,du. \tag{12}$$

This general method of obtaining an antiderivative of $f(x)$ explains the more usual name and symbol for it. The name is *an indefinite integral of* $f(x)$; the symbol is $\int f(x)\,dx$, without indicating the limits of integration.

The arithmetic process implied in the integration of a function of one independent variable is typical of all integration: ordinary summation leads to a sequence the limit of which is an integral — it may be an ordinary definite integral or a line integral or a surface integral or a volume integral or an integral of a function of a complex variable.

If x is time and $f(x)$ is the speed of a particle, then the integral (8) is the distance traversed by the particle between two instants. If $f(x)$ is the rate of flow of water from a pipe, the integral is the quantity of water which has

left the pipe in the time interval (a,x). Either picture is representative of the definite integral of a function of one independent variable.

Problems

1. Show that

$$\int \cos x\,dx = \sin x + C, \qquad \int \sec^2 x\,dx = \tan x + C,$$

$$\int \tan^2 x\,dx = \tan x - x + C, \qquad \int \sin^2 x\,dx = \tfrac{1}{2}x - \tfrac{1}{4}\sin 2x + C,$$

$$\int \cos^2 x\,dx = \tfrac{1}{2}x + \tfrac{1}{4}\sin 2x + C,$$

$$\int \sqrt{1-x^2}\,dx = \tfrac{1}{2}x\sqrt{1-x^2} + \tfrac{1}{2}\sin^{-1} x + C.$$

2. Let $M(x)$ be a function whose derivative is $1/x$. Show that

$$\int \tan x\,dx = -M(\cos x) + C, \qquad \int \cot x\,dx = M(\sin x) + C,$$

$$\int \frac{dx}{\sin x} = M\left(\tan\frac{x}{2}\right) + C,$$

$$\int \frac{dx}{1-x^2} = \frac{1}{2}\int\left(\frac{1}{1+x} + \frac{1}{1-x}\right)dx = \tfrac{1}{2}M(1+x) - \tfrac{1}{2}M(1-x) + C,$$

$$\int \frac{dx}{\sqrt{1+x^2}} = M(x + \sqrt{1+x^2}) + C.$$

3. Show that $\int u(x)\,dv(x) = u(x)v(x) - \int v(x)\,du(x)$.

4. Show that if $M(x)$ is defined as in Problem 2, then

$$\int M(x)\,dx = xM(x) - x + C.$$

5. Show that

$$\int x\cos x\,dx = x\sin x + \cos x + C,$$

$$\int x^2\cos x\,dx = x^2\sin x + 2x\cos x - 2\sin x + C.$$

6. Show that, if $M(x)$ and $N(x)$ are functions whose derivatives are given below,

$$\frac{d}{dx}M(x) = \frac{\sin x}{x}, \qquad \frac{d}{dx}N(x) = \frac{\cos x}{x},$$

then $$\int \frac{\sin x}{x}\,dx = M(x) + C, \qquad \int \frac{\cos x}{x}\,dx = N(x) + C,$$

$$\int \frac{\sin x}{x^2}\,dx = -\frac{\sin x}{x} + N(x) + C, \qquad \int \frac{\cos x}{x^2}\,dx = -\frac{\cos x}{x} - M(x) + C,$$

$$\int \frac{\sin x}{x^3}\,dx = -\frac{\sin x}{2x^2} - \frac{\cos x}{2x} - \frac{1}{2}M(x) + C.$$

2. *Area under a curve as a geometric representation of an integral*

Figure 6.1 presents one possible geometric representation of the solution of the differential equation (2). Another representation is given in Figure 6.2 where we begin with the curve representing the *given* function $u = f(x)$.

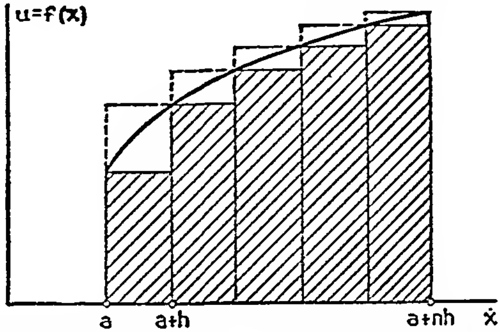

Fig. 6.2. The definite integral of a given function $u = f(x)$ may be interpreted as the area under the curve $u = f(x)$.

The successive terms of the sum (8) are the areas of the shaded rectangles and the limit of the sum is the area under the curve. The picture suggests that

$$\int_a^x f(x)\,dx = \lim \sum_{m=1}^{n} f(a + mh)h, \tag{13}$$

since the area under the curve is also the limit of the sum of the areas of the augmented rectangles. If in the interval under consideration, $f(x)$ is a *monotonic* function (either a nonincreasing or a nondecreasing function), the value of the integral lies between the sums in equations (8) and (13) and the error of approximating the integral by either of the two sums is less than $|f(x) - f(a)|h$.

The average of the two sums is an even better approximation to the integral

$$\int_a^x f(x)\,dx \simeq [\tfrac{1}{2}f(a) + f(a + h) + f(a + 2h) + \cdots$$
$$+ f(a + \overline{n-1}h) + \tfrac{1}{2}f(a + nh)]h, \tag{14}$$

where $x = nh$. Geometrically, this means adding the areas of trapezoids as in Figure 6.3. The error is less than $\frac{1}{2}|f(x) - f(a)|h$.

The following table illustrates the process of numerical integration

x	1.0	1.1	1.2	1.3	1.4	1.5
$\cos x$	0.540	0 454	0 362	0 268	0.170	0.071
$\int_1^x \cos x\, dx$	0.0	0.0497	0.0905	0.1220	0.1439	0.1559

The first entry in the last row is zero. The next entry is the average of the first two entries in the second row multiplied by the length 0.1 of the interval. One tenth of the average of the second and third entries in the second row is added to obtain the next entry in the last row; and the process is thus continued. Since the antiderivative is $\sin x$, the integral is $\sin x - \sin 1$. If $x = 1.5$, this quantity is 0.15602. Considering that we have retained only three decimal places in the first row, the accumulated error of one unit in the fourth place of the last entry is not excessive.

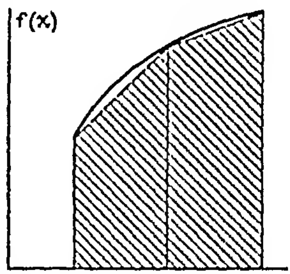

FIG. 6.3. Illustrating the trapezoidal rule (14) for approximating a definite integral.

Naturally, throughout the entire discussion it has been assumed that the processes under consideration could be carried out and that the given functions could be called *integrable*. "Integrable in the Riemann sense" we should have said; for there is no limit to man's ingenuity and there are integrals in the sense of Lebesgue, Stieltjes, etc. It is easy to show that there are functions which are not integrable in the Riemann sense. If, for example, $f(x) = +1$ or -1, depending on whether x is rational or irrational, it would be impossible to carry out the above directions for integration.

Problems

1. Prepare a table of $\int_1^x \frac{dx}{x}$ in the interval (1,2) and compare it with a table of $\log x$ (natural logarithm).

2. The *integral sine*

$$\mathrm{Si}\, x = \int_0^x \frac{\sin x}{x}\, dx$$

occurs in the theory of radiation. Prepare a short table of $\mathrm{Si}\, x$ and of

$$\int_0^x \mathrm{Si}\, x\, dx.$$

3. Prepare a table of $\int_0^x I_0(x)\,dx$, where $I_0(x)$ is the modified Bessel function of order zero and of the first kind.

4. Using the quadratic interpolation formula derive the following approximation

$$\int_{x_0}^{x_0+nh} y\,dx = \tfrac{1}{3}h(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + 4y_5 + 2y_6 + \cdots + 4y_{n-1} + y_n),$$

where $y_m = y(x_0 + mh)$ and n is even. This expression is the sum

$$\tfrac{1}{3}h(y_0 + 4y_1 + y_2) + \tfrac{1}{3}h(y_2 + 4y_3 + y_4) + \tfrac{1}{3}h(y_4 + 4y_5 + y_6) + \cdots.$$

If $n = 2m + 1$, the value of the integral is obtained for $n = 2m$; to this the following quantity is added: $\frac{1}{2}(y_{2m} + y_{2m+1})h + \frac{1}{12}h(2y_{2m+1} - y_{2m} - y_{2m+2})$. The second term is an improvement on linear interpolation.

The above approximation to the integral is known as *Simpson's rule.*

5. Obtain an approximate formula based on cubic interpolation.

6. Show analytically and illustrate graphically that

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(t,x)\,dx = \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(t,x)\,dx + f(t,b)\frac{db}{dt} - f(t,a)\frac{da}{dt}.$$

7. Show that, if a and b are constant and $f(t,x)$ has a finite discontinuity at $x = \xi$ where ξ is independent of t,

$$\frac{d}{dt}\int_a^b f(t,x)\,dx = \int_a^b \frac{\partial}{\partial t} f(t,x)\,dx.$$

8. Show that if the discontinuity of $f(t,x)$ is at $x = t$, where $a < t < b$, the equation of Problem 7 must be replaced by

$$\frac{d}{dt}\int_a^b f(t,x)\,dx = \int_a^b \frac{\partial}{\partial t} f(t,x)\,dx + f(t,t-0) - f(t,t+0).$$

9. Let

$$\begin{aligned} f(t,x) &= (1-t)x, \qquad 0 \le x < t, \\ &= t(x-1), \qquad t < x \le 1. \end{aligned}$$

Evaluate

$$I(t) = \int_0^1 f(t,x)\,dx,$$

show that $dI/dt = -3t^2 + 3t - 0.5$, and thus verify the formula in Problem 8.

10. Let $f(t,x) = tx$. Evaluate

$$I(t) = \int_t^{t^2} f(t,x)\,dx,$$

obtain the derivative of $I(t)$, and thus verify the formula in Problem 6.

3. *Line integral*

In an electromagnetic field the voltage, or electromotive force, between infinitely close points (x,y,z) and $(x + \Delta x, y + \Delta y, z + \Delta z)$ is of the following form

$$P(x,y,z)\Delta x + Q(x,y,z)\Delta y + R(x,y,z)\Delta z + \text{infinitesimals of higher order.}$$

Functions P, Q, R are the components of the electric intensity in the x, y, z directions. Proceeding by infinitesimal steps along some curve joining two given points, A and B, we have an expression for the total electromotive force

$$V = \int_{AB} (P\,dx + Q\,dy + R\,dz) = \lim \sum (P_m\,\Delta x_m + Q_m\,\Delta y_m + R_m\,\Delta z_m). \quad (15)$$

In the summation P_m, Q_m, R_m refer to some point within the mth interval on the curve AB. Integrals of this type are called *line integrals.*

An ordinary definite integral is a special line integral: the curve AB is a segment of the x-axis and the integrand depends only on x. The evaluation of line integrals can, at least in principle, be reduced to ordinary integration. If

$$V = \int_{AB} (y\,dx - x\,dy) \quad (16)$$

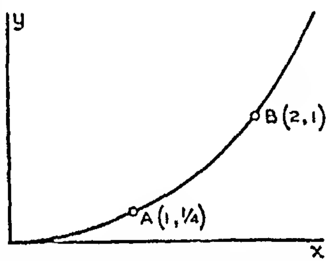

FIG. 6.4. A segment AB of the parabola $y = \frac{1}{4}x^2$ along which the line integral (16) is to be evaluated.

is taken along the parabola $y = \frac{1}{4}x^2$ from $(1,1/4)$ to $(2,1)$ (see Figure 6.4), we know in effect a relationship between the differentials, $dy = \frac{1}{2}x\,dx$. Substituting in (16), we have

$$V = \int_1^2 (\tfrac{1}{4}x^2\,dx - \tfrac{1}{2}x^2\,dx) = -\tfrac{1}{12}x^3\Big|_1^2 = -7/12.$$

If the relation between x and y is implicit, $f(x,y) = 0$, it may be necessary to depend on the definition (15) to evaluate the integral.

It is clear that in general the line integral depends on the path of integration and not merely on the end points. If, however, the integrand is an exact differential of some function, then the line integral becomes independent of the path. For instance, for any curve connecting two points $A(x_1,y_1)$ and $B(x_2,y_2)$

$$V = \int_{AB} (x\,dx + y\,dy) = \tfrac{1}{2}(x^2 + y^2)\Big|_{AB} = \tfrac{1}{2}(x_2^2 + y_2^2) - \tfrac{1}{2}(x_1^2 + y_1^2). \quad (17)$$

In this case the element $x\,\Delta x + y\,\Delta y$ of the sum (15) differs from $\Delta(\frac{1}{2}x^2 + \frac{1}{2}y^2)$ by infinitesimals of higher order and the sum of successive increments of the expression $\frac{1}{2}x^2 + \frac{1}{2}y^2$ is evidently the difference between its final and initial values.

In the above example the integrand is the differential of a single-valued function. In the case of a multiple-valued function care should be taken to take the proper values at the end points. For example,

$$\varphi = \int_{AB} \frac{x\,dy - y\,dx}{x^2 + y^2} = \int_{AB} d\tan^{-1}(y/x)$$
$$= \tan^{-1}(y_2/x_2) - \tan^{-1}(y_1/x_1). \quad (18)$$

The integrand is a single-valued function; and if AB does not go through the origin, there is nothing ambiguous about it. On the other hand, the antidifferential $\tan^{-1}(y/x)$ is a multiple-valued function; in Figure 6.5 this function is represented by the angle φ which can assume either its smallest value or any value differing from it by an integral multiple of 2π. Thus the final answer in (18) is ambiguous. This ambiguity is resolved if we note that the successive increments in the sum (15) are infinitesimal and that in computing $\Delta\tan^{-1}(y/x)$ we should take the difference between two nearby values of the inverse tangent; this difference is free from ambiguity. Similarly there is no ambiguity in the difference between the values at the end points of a curve *if we follow the curve.*

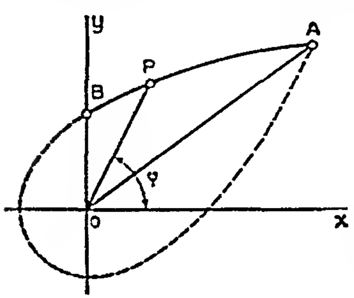

FIG. 6.5. The angle φ represents the antidifferential of the function $(x\,dy - y\,dx)/(x^2 + y^2)$.

To illustrate, let us consider a path of integration connecting $A(1,1)$ with $B(0,0.5)$, as shown by the solid curve in Figure 6.5. If we start with the initial value of $\varphi = \tan^{-1}(y/x) = \pi/4$ at A, the corresponding value at B will be $\pi/2$ and the value of the integral will be $\pi/4$. If we start with $\varphi = 9\pi/4$, we end up with $10\pi/4$; the difference is still $\pi/4$. If, however, we follow the dotted path, φ decreases from $\pi/4$ to zero, becomes negative and reaches the value $-3\pi/2$ at B; hence, the line integral along the dotted path is $(-3\pi/2)-(\pi/4) = -7\pi/4$.

With regard to the dependence of the line integral in a plane on the path of integration *connecting two fixed points* we have encountered three possibilities:

(a) In general, the line integral varies with a variation in the path, although there may be some variations for which the line integral remains constant. Thus we could deform one part of the path and, having de-

termined the change in the line integral, we could deliberately alter the remainder of the path to compensate for this change.

(b) If the integrand is a total differential of a single-valued function, the line integral is entirely independent of the path.

(c) If the integrand is a total differential of a multiple-valued function, the line integral is independent of path deformation so long as the path does not cross certain points (the origin in the above example). In crossing such "singular points" the line integral varies discontinuously.

Another way of stating these properties is:

(a) Generally, the line integral round a closed curve is different from zero.

(b) If the integrand is the total differential of a single-valued function, the line integral round any closed curve is zero.

(c) If the integrand is the total differential of a multiple-valued function, the line integral round a closed curve does not change with any deformation of the curve which does not involve crossing certain singular points, or, in the case of three variables, singular lines.

There are examples of physical fields of force falling into each of these three categories. Suppose that E_x, E_y, E_z are the cartesian components of the electric intensity in an electrostatic field; then $E_x\,dx + E_y\,dy + E_z\,dz$ is the exact differential of a certain single-valued "potential function" and the line integral round any closed curve is zero. This line integral represents the work done by the field on a unit charge; hence, the electrostatic field does no work on a charge which has been moved completely round a closed curve. Such fields of force are called *conservative.*

If H_x, H_y, H_z are the components of the magnetic intensity of the field generated by steady electric currents in a number of closed loops, then $H_x\,dx + H_y\,dy + H_z\,dz$ is the exact differential of a multiple-valued function. The line integral is the magnetomotive force and it is independent of any deformation in the path which does not involve crossing any of the current-carrying loops. Whenever the path crosses a particular loop, the magnetomotive force is either augmented or diminished by an amount equal to the current in the loop.

Finally, in the case of a nonsteady electromagnetic field, neither of the above differentials is exact and the line integrals depend on the paths of integration.

Problems

1. Calculate (16) first along the parabola $y = x^2$ from (0,0) to (1,1) and then along $y^2 = x$, between the same points. Show that the difference is twice the area bounded by the two curves.

2. Evaluate $V = \int (z\,dx + x^2\,dy - 2y\,dz)$, starting from the origin along the curve given by the following parametric equations $x = t, y = t^2, z = t^3$.
Ans. $\frac{3}{4}t^4 - \frac{6}{5}t^5$.

3. Show that the line integral (18) taken once in the counterclockwise direction round a circle concentric with the origin is 2π. Do this without making direct use of the fact that the integrand is the differential of the inverse tangent. *Hint:* Use a certain parametric equation of the circle.

4. *Surface integral*

In a *surface integral* the integration is extended over an area in a plane or, more generally, over a surface in space. As a representative picture of the surface integral we may take the flow of liquid across a surface. The flow of liquid in a given interval of time across an element of surface ΔS becomes more nearly proportional to ΔS as ΔS approaches zero and the coefficient of proportionality $f(x,y,z)$ is the flow per unit area. The total flow across the surface is the integral

$$\int\int f(x,y,z)\,dS = \lim_{\Delta S_m \to 0} \sum f(x_m,y_m,z_m)\,\Delta S_m. \tag{19}$$

If the surface of integration is a plane, a convenient partition into elements is accomplished with a net of straight lines parallel to the coordinate axes, Figure 6.6, in which case $dS = dx\,dy$. Any method of subdivision

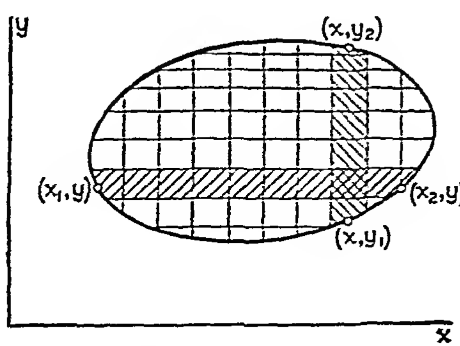

FIG. 6.6. A partition of an area in the definition of a surface integral.

and any order of summation should yield the same limit; otherwise the surface integral would not exist in the sense implied by (19) and by the picture of liquid flow. It is possible to define functions which are *not* integrable in this sense; but such functions are of no interest in applied mathematics.

A simple example of a surface integral is the volume bounded by a

surface $z = f(x,y)$, the xy-plane, and a cylindrical surface whose generators are parallel to the z-axis. This volume is the surface integral, $\int\int z\,dx\,dy$, taken over the base of the cylinder in the xy-plane; the integrand $z\,dx\,dy$ is the volume of an elementary rod parallel to the z-axis.

If $f(x,y,z)$ is the density of electric charge on the surface of a conductor, that is, the charge per unit area, then (19) represents the total charge on the conductor or on a specified part of it. The electric potential of a point charge equals the charge divided by $4\pi\epsilon r$, where ϵ is the dielectric constant and r is the distance of the point charge from a representative point in space. The potential of the surface distribution is then the following surface integral

$$\int\int \frac{f(x,y,z)\,dS}{4\pi\epsilon\sqrt{(x-\hat{x})^2+(y-\hat{y})^2+(z-\hat{z})^2}}, \tag{20}$$

where $(\hat{x},\hat{y},\hat{z})$ is a point in space and the integration is extended over the surface of the conductor.

5. *Volume integral*

If $f(x,y,z)$ is the density of a body, the total mass is the *volume integral*

$$\int\int\int f(x,y,z)\,d\tau = \lim_{\Delta\tau\to 0} \sum f(x_m,y_m,z_m)\,\Delta\tau_m, \tag{21}$$

where $d\tau$ is a differential element of volume.

The same integral would express the total electric charge in a given volume provided $f(x,y,z)$ is the density of charge. If dS in (20) is replaced by $d\tau$, the integral will represent the electric potential of a given volume distribution of charge.

6. *Integral of a function of a complex variable*

The complex variable is defined in a plane and the integral

$$\int f(z)\,dz = \lim_{\Delta z_m\to 0} \sum f(z_m)\,\Delta z_m \tag{22}$$

is essentially a line integral. Substituting $f(z) = u(x,y) + iv(x,y)$ and $dz = dx + i\,dy$, we have

$$\begin{aligned}\int f(z)\,dz &= \int [u(x,y) + iv(x,y)]\,(dx + i\,dy) \\ &= \int (u\,dx - v\,dy) + i\int (v\,dx + u\,dy).\end{aligned} \tag{23}$$

The first of the above equations seems to suggest that the integral depends only on the end points of the path of integration; the second indicates that it may depend on the path as well. What actually happens depends on whether $f(z)$ is obtained by performing arithmetic operations on z as a whole or on its real and imaginary parts separately, that is, on whether the integrand is monogenic or polygenic. The difference may be illustrated by the following examples. Thus

$$\int_{z_1}^{z_2} z\,dz = \tfrac{1}{2}z^2\Big|_{z_1}^{z_2} = \tfrac{1}{2}z_2^2 - \tfrac{1}{2}z_1^2 = \int (x+iy)(dx+i\,dy) \tag{24}$$

depends only on the end points. Just as in the case of a real variable, $z\,\Delta z = \Delta(z^2/2)$ except for infinitesimals of higher order. The summation along the curve reduces to the addition of successive differences to $z_1^2/2$ and the result is simply the difference between the final and initial values.

On the other hand, the integral

$$\int (x-iy)(dx+i\,dy) = \int (x\,dx+y\,dy) + i\int (x\,dy - y\,dx) \tag{25}$$

depends on the actual path of integration. The function $x - iy$ can be obtained from $x + iy$ only if x and y are considered as separate entities. In this case there is no antiderivative for the entire integrand. It happens that the first integrand is an exact differential and the integral equals the difference of the final and initial values of $\frac{1}{2}(x^2+y^2)$; but the second integral depends on the actual path. Integrating along the x-axis from $z = 0$ to $z = 1$ where $y = 0$ and $dy = 0$, and then along a line parallel to the y-axis to point $z = 1 + i$ where $x = 1$ and $dx = 0$, we obtain $1 + i$. Integrating along the y-axis and then along the line parallel to the x-axis, we obtain $1 - i$. Following the first path and coming back to $z = 0$ by the second, that is, integrating round the boundary of the unit square, we get $2i$.

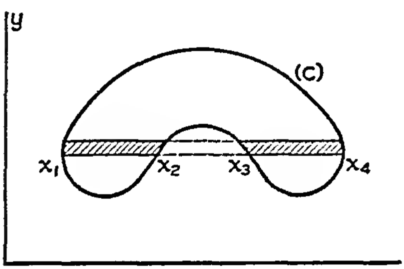

FIG. 6.7. An area of integration.

7. *Green's theorem*

Green's theorem establishes a relation between surface and line integrals. We shall consider a simple case first. The integral

$$A = \int\int dx\,dy \tag{26}$$

represents the area of integration, Figure 6.7; this simply states that the area is the sum of its elements. Integrating with respect to x, we obtain

$$A = \int (x_2 - x_1)\,dy. \tag{27}$$

We note that on the right side of the boundary (C) of the area, the positive differential element dy corresponds to a counterclockwise movement on (C); on the left side, the positive dy corresponds to a clockwise and $-dy$ to a counterclockwise movement. Therefore, the above integral represents a line integral.

$$A = \int_C x\,dy, \tag{28}$$

taken counterclockwise.

Similarly,

$$A = -\int_C y\,dx, \tag{29}$$

also taken counterclockwise. From these equations we have

$$A = \tfrac{1}{2}\int_C (x\,dy - y\,dx). \tag{30}$$

Thus the surface integral (26) has been expressed as a line integral.

We shall now show that a more general line integral

$$S = \int_C (P\,dx + Q\,dy) \tag{31}$$

is equal to the surface integral

$$S = \int\int\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy. \tag{32}$$

Integrating the first term of (32) with respect to x, we have

$$\begin{aligned}\int\int \frac{\partial Q}{\partial x}\,dx\,dy &= \int [Q(x_2,y) - Q(x_1,y)]\,dy \\ &= \int_C Q(x,y)\,dy.\end{aligned} \tag{33}$$

Similarly, integrating the second term of (32) with respect to y, we obtain the first term of (31).

The extension of Green's theorem to curved surfaces and an analogous relation between volume and surface integrals are considered in the next chapter.

An immediate corollary of Green's theorem is a condition under which the line integral round a closed curve vanishes. If

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y} \tag{34}$$

is true *everywhere* in the region bounded by (C), the line integral (31) will certainly vanish. It should be noted that the line integral depends only on the values of P and Q on the boundary (C), and may exist when the surface integral does not.

Simple examples will show that, unless the condition (34) is satisfied throughout the area *in spirit as well as in form*, we cannot be sure that (31) vanishes without further investigation. Consider the integrals

$$S_1 = \int_C \frac{x\,dx + y\,dy}{x^2 + y^2}, \qquad S_2 = \int \frac{x\,dy - y\,dx}{x^2 + y^2}, \tag{35}$$

taken round a circle concentric with the origin at which point neither P nor Q exists. The derivatives,

$$\begin{aligned} \frac{\partial Q_1}{\partial x} &= -\frac{2xy}{(x^2 + y^2)^2}, & \frac{\partial P_1}{\partial y} &= -\frac{2xy}{(x^2 + y^2)^2}; \\ \frac{\partial Q_2}{\partial x} &= \frac{y^2 - x^2}{(x^2 + y^2)^2}, & \frac{\partial P_2}{\partial y} &= \frac{y^2 - x^2}{(x^2 + y^2)^2}, \end{aligned} \tag{36}$$

satisfy (34) formally; but they do not exist at the origin. Hence, the separate parts of the surface integral do not exist, Green's theorem is not applicable, and no conclusion about the values of S_1 and S_2 can be drawn without further study. Let $x = a \cos \varphi$ and $y = a \sin \varphi$, where a is the radius of the circle of integration. Integrating from $\varphi = 0$ to $\varphi = 2\pi$, we find that $S_1 = 0$ and $S_2 = 2\pi$. Thus (34) is a sufficient but not a necessary condition for the vanishing of (31).

8. *Evaluation of integrals*

Line, surface and volume integrals are almost indispensable in the formulation of many physical laws. Thus, in one form, Maxwell's equations represent the equality of certain line and surface integrals. In such circumstances one need know only the meaning of various kinds of integrals. Then comes a time when certain integrals have to be evaluated. In principle all integrals can be reduced to ordinary definite integrals; but in practice it may be more expedient to resort to numerical evaluation based on the definitions. We have seen that a line integral becomes an ordinary definite integral if all the variables can be expressed in terms of any one variable. Suppose, however, the equation of the path of integration, $f(x,y) = 0$, is given in an implicit form; unless this equation can be solved analytically either for x or for y, or both can be expressed in terms of some parameter t, the simplest method of evaluation is the one inherent in the definition of the line integral. The same is true of surface and volume integrals.

In this section we shall consider a few examples in which integration can be carried out without resorting to numerical methods. Let the surface integral

$$S = \int\int xy\,dx\,dy \tag{37}$$

be taken over the area bounded by the straight line, $y = x$, and the parabola, $y = x^2$, Figure 6.8. We may integrate first with respect to x, thus adding the elements in a strip parallel to the x-axis where y is constant. The limits for this integration are $x = y$ and $x = \sqrt{y}$; thus

$$\begin{aligned} S &= \int y\,dy \int_y^{\sqrt{y}} x\,dx = \int y\,dy\,(\tfrac{1}{2}x^2)\Big|_y^{\sqrt{y}} \\ &= \tfrac{1}{2}\int y\,dy(y - y^2) = \tfrac{1}{2}\int (y^2 - y^3)\,dy. \end{aligned} \tag{38}$$

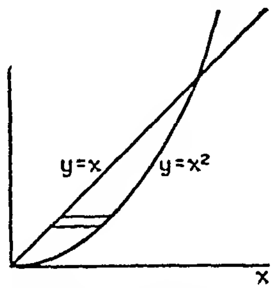

FIG. 6.8. Showing the area of integration for the surface integral (37).

Next we integrate the strips over the area of integration where y varies from zero to unity (the second point of intersection of the boundary curves) and obtain

$$S = \tfrac{1}{2}(\tfrac{1}{3}y^3 - \tfrac{1}{4}y^4)\Big|_0^1 = \tfrac{1}{24}.$$

Integrating first with respect to y and then with respect to x, we have

$$S = \int_0^1 x\,dx \int_{x^2}^{x} y\,dy = \tfrac{1}{2}\int_0^1 x\,dx(y^2)\Big|_{x^2}^{x} = \tfrac{1}{2}\int_0^1 (x^3 - x^5)\,dx = \tfrac{1}{24}. \tag{39}$$

The principal complication is in the variable limits; otherwise each step consists of ordinary integration. Sometimes one order of integration is simpler to carry out than the other. At times some net other than rectangular is preferable. Consider, for example,

$$S = \int\int (x^2 + y^2)\,dx\,dy \tag{40}$$

taken over a circle of radius a with the center at the origin The polar coordinate system suggests itself immediately, for $x^2 + y^2 = \rho^2$ and the limits of integration with respect to ρ and φ are constant. The element of area in polar coordinates is $\rho\,d\rho\,d\varphi$ and (40) becomes

$$S = \int_0^{2\pi}\int_0^a \rho^3\,d\rho\,d\varphi = \left(\int_0^a \rho^3\,d\rho\right)\left(\int_0^{2\pi} d\varphi\right) = \tfrac{1}{2}\pi a^4. \tag{41}$$

The meaning of the integral is the guide in transforming from (40) to

(41). It would not be right merely to replace x and y by their expressions in polar coordinates. The element of area, $dx\,dy$, is proper for the cartesian subdivision of the total area by straight lines parallel to the coordinate axes, but not for repeated integration in polar coordinates. For a constant φ, the integration has to be performed in an angular strip and the rectangular elements do not fit into it. In changing coordinates in a surface integral, the differential element of area bounded by coordinate lines in one system must be replaced by the differential element of area bounded by coordinate lines in the second system; thus, $dx\,dy$ is replaced by $\rho\,d\rho\,d\varphi$, Figure 6.9, and $d\rho\,d\varphi$ by $(dx\,dy)/\rho$. The following equations show the difference between this transformation and mere substitution of variables

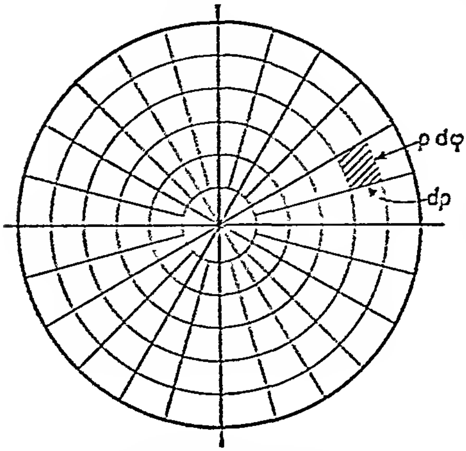

FIG. 6.9. A partition appropriate for the evaluation of surface integrals in polar coordinates.

$$\begin{gathered} dx = \cos\varphi\,d\rho - \rho\sin\varphi\,d\varphi, \qquad dy = \sin\varphi\,d\rho + \rho\cos\varphi\,d\varphi, \\ dx\,dy = \sin\varphi\cos\varphi\,d\rho^2 + \rho(\cos^2\varphi - \sin^2\varphi)\,d\rho\,d\varphi - \rho^2\sin\varphi\cos\varphi\,d\varphi^2. \end{gathered} \tag{42}$$

Transformation of coordinates is sometimes useful in evaluating definite integrals. Consider

$$S = \int_0^\infty e^{-x^2}\,dx. \tag{43}$$

The variable x is a dummy variable and we can write equally well

$$S = \int_0^\infty e^{-y^2}\,dy. \tag{44}$$

Multiplying (43) and (44) and transforming to polar coordinates, we have

$$\begin{aligned} S^2 &= \int_0^\infty e^{-x^2}\,dx \int_0^\infty e^{-y^2}\,dy = \int_0^\infty \int_0^\infty e^{-x^2-y^2}\,dx\,dy \\ &= \int_0^{\pi/2}\int_0^\infty e^{-\rho^2}\rho\,d\rho\,d\varphi = \left(\int_0^\infty e^{-\rho^2}\rho\,d\rho\right)\int_0^{\pi/2} d\varphi \\ &= \left(-\tfrac{1}{2}e^{-\rho^2}\right)\Big|_0^\infty (\varphi)\Big|_0^{\pi/2} = \pi/4. \end{aligned} \tag{45}$$

Therefore, $S = \sqrt{\pi}/2$.

We have treated the integrals taken over an infinite range as if they were sums of a finite number of terms. This is a safe procedure so long as the integrand decreases rapidly as the variable of integration increases. When the integrals are only just convergent the operation of multiplication should be avoided, at least without consulting the more advanced treatises where such borderline cases are studied. Examples of such integrals are given in the next section.

In the case of volume integrals there is one more integration to perform. Suppose that we wish to find the total mass of a hemisphere of radius a when the density is proportional to the distance from the plane surface. If this surface is chosen as the xy-plane and the density is assumed equal to kz, then the mass is given by

$$m = k \int\int\int z\, dx\, dy\, dz. \tag{46}$$

The integrand is dependent on z alone and the quickest method is to integrate first with respect to x and y

$$m = k \int z\, dz \int\int dx\, dy. \tag{47}$$

Since the equation of the sphere is

$$x^2 + y^2 + z^2 = a^2, \tag{48}$$

the boundary of the area of integration is the circle $x^2 + y^2 = a^2 - z^2$. In this case the surface integral is, in fact, the area of the circle and therefore

$$m = \pi k \int_0^a z(a^2 - z^2)\, dz = \tfrac{1}{4}\pi k a^4. \tag{49}$$

Suppose, however, that we integrate first with respect to z; then

$$m = k \int\int dx\, dy \int_0^{\sqrt{a^2-x^2-y^2}} z\, dz = \tfrac{1}{2}k \int\int (a^2 - x^2 - y^2)\, dx\, dy.$$

The variables x and y can assume any value within the base of the hemisphere and this base is bounded by the circle $x^2 + y^2 = a^2$. Since x varies from $-\sqrt{a^2 - y^2}$ to $\sqrt{a^2 - y^2}$, we have

$$\begin{aligned} m &= \tfrac{1}{2}k \int dy \int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} (a^2 - x^2 - y^2)\, dx \\ &= \tfrac{1}{2}k \int dy\, (a^2x - \tfrac{1}{3}x^3 - y^2x)\Big|_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} \\ &= \tfrac{2}{3}k \int (a^2 - y^2)\sqrt{a^2 - y^2}\, dy. \end{aligned}$$

The limits of the final integration are $-a$, $+a$. Introducing a new variable t such that $y = a \sin t$, we obtain

$$m = \tfrac{2}{3}ka^4 \int_{-\pi/2}^{\pi/2} \cos^4 t\, dt = \tfrac{4}{3}ka^4 \int_0^{\pi/2} \cos^4 t\, dt.$$

If the integrand is expressed in terms of multiple angles (see Section 1.7),

$$m = \tfrac{1}{6}ka^4 \int_0^{\pi/2} (3 + 4\cos 2t + \cos 4t)\, dt = \tfrac{1}{4}\pi ka^4.$$

This long laborious step by step integration is the rule; simplifications may sometimes be spectacular but nevertheless they are exceptions depending on the character of the integrand and on the region of integration.

For further information on the evaluation of integrals the reader is referred to Joseph Edwards, "A Treatise on the Integral Calculus," Vol. II (Macmillan and Company).

9. *Improper and infinite integrals*

An *improper integral* is an integral in which the integrand becomes infinite within the range of integration. Thus

$$I = \int_0^1 \frac{dx}{\sqrt{x}} \tag{50}$$

is an improper integral. If a small region containing the singularity of the integrand is isolated, the rest of the integral becomes proper and the improper integral is regarded as the limit of the proper integral as the isolated region is made to approach zero. Thus

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim \int_\delta^1 \frac{dx}{\sqrt{x}} \qquad \text{as } \delta \to 0. \tag{51}$$

The "convergence" of the improper integral depends solely on the behavior of the integrand in the immediate vicinity of the singularity. In the above case the antiderivative is $2\sqrt{x}$ and the proper integral is equal to $(2 - 2\sqrt{\delta})$; as δ approaches zero the limit is 2. If the integrand were $1/x$, the antiderivative would be $\log x$ and the improper integral would diverge.

An *infinite integral* is an integral taken over an infinite range as in

$$I = \int_1^\infty \frac{dx}{x^2} = \lim \int_1^A \frac{dx}{x^2} \qquad \text{as } A \to \infty. \tag{52}$$

Thus the infinite integral is regarded as the limit of a proper integral in which the range of integration is allowed to increase indefinitely. If the proper integral can be evaluated, the question of convergence can readily be settled and the value of the infinite integral found by direct passage to the limit. Thus in (52) the proper integral is $(1 - 1/A)$; as A increases, this value approaches unity. If the integrand were $1/x$, the proper integral would be $\log A$ and the infinite integral would diverge.

If we are unable to obtain the proper integral in a sufficiently simple form, we can still settle the question of convergence by studying the behavior of the integrand. Just as in the case of infinite series, the comparison test is the most important test for the convergence of an integral. Thus, if we multiply the integrand in (52) by 10^{-x}, we can be sure that the resulting integral will converge even though we may find it difficult to evaluate.

There is a connection between the convergence properties of infinite integrals and infinite series. Thus if $f(x)$ is a monotone (nonincreasing or nondecreasing) function of x, then

$$\int_1^\infty f(x)\,dx, \qquad \sum_1^\infty f(n) \tag{53}$$

converge or diverge together. This is clear from Figure 6.10 where the area under the curve is the integral, the sum of the areas of the rectangles projecting outside the curve is the sum of the series, and the sum of the areas of the rectangles which do not project outside the curve is the sum of the series diminished by $f(1)$.

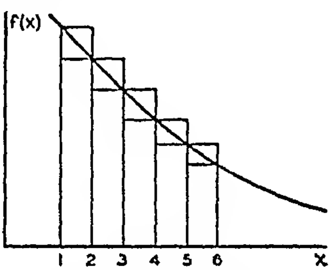

Fig. 6.10. Illustrating the relationship between the convergence of infinite integrals and infinite series.

Sometimes a transformation of the variable of integration helps to establish the convergence of an infinite integral. Take the following integrals encountered in the theory of the diffraction of waves

$$A = \int_0^\infty \cos x^2\,dx, \qquad B = \int_0^\infty \sin x^2\,dx. \tag{54}$$

As x increases, the integrands fluctuate between $+1$ and -1. These fluctuations become more rapid and the areas under positive half-waves and over negative diminish. Thus the integrals should converge. We can make this conclusion clearer, however, if we let $x^2 = t$, $2x\,dx = dt$;

$$A = \tfrac{1}{2}\int_0^\infty \frac{\cos t}{\sqrt{t}}\,dt, \qquad B = \tfrac{1}{2}\int_0^\infty \frac{\sin t}{\sqrt{t}}\,dt. \tag{55}$$

The infinite series

$$A = \frac{1}{2}\left(\int_0^{\pi/2} \frac{\cos t}{\sqrt{t}}\,dt + \int_{\pi/2}^{3\pi/2} \frac{\cos t}{\sqrt{t}}\,dt + \int_{3\pi/2}^{5\pi/2} \frac{\cos t}{\sqrt{t}}\,dt + \cdots\right),$$

$$B = \frac{1}{2}\left(\int_0^{\pi} \frac{\sin t}{\sqrt{t}}\,dt + \int_{\pi}^{2\pi} \frac{\sin t}{\sqrt{t}}\,dt + \int_{2\pi}^{3\pi} \frac{\sin t}{\sqrt{t}}\,dt + \cdots\right),$$

are alternating series with the nth terms approaching zero. Hence the limits exist.

CHAPTER VII

VECTOR ANALYSIS

1. *Scalars and vectors*

A large number of physical quantities are grouped together under the generic name *scalars;* numerous other quantities are called *vectors.* Scalars may be vastly different in their physical nature but quantitatively they are characterized by a single number expressing the "magnitude" of the quantity. Mass, volume, weight, distance between two points, an interval of time, electric potential, stiffness of a spring — are all scalars. Addition of scalars of the same physical nature is just ordinary arithmetic addition.

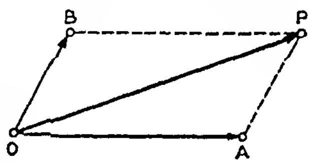

FIG. 7.1. The sum OP of two vectors OA and OB is defined, both in magnitude and direction, as that diagonal of the parallelogram constructed on OA and OB which emerges from the common point O of the two vectors.

Force, velocity, acceleration are vectors. To characterize them it is necessary to specify their direction as well as their magnitude. The law of addition is not that of ordinary arithmetic. The magnitude of the resultant of two equal forces, acting perpendicularly to each other, is not twice as great as either force but only $\sqrt{2}$ times as great; its direction is that of the bisector of the 90° angle formed by the two forces. To find the resultant of any two forces acting at the same point we draw radii in the direction of each of the forces, with their lengths proportional to the magnitudes of the forces, and construct a parallelogram, Figure 7.1. The diagonal OP, emerging from the common origin of the radii OA and OB which represent the forces, gives the magnitude and relative direction of the resultant force. This is the parallelogram law of addition characteristic of all vectors; it represents a test for deciding whether or not a given quantity is a vector.

Symbolically we may write

$$\overrightarrow{OA} + \overrightarrow{OB} = \overrightarrow{OP}, \tag{1}$$

where the arrows indicate the vectorial character of the segments as distinct from mere lengths OA, OB, OP. If, however, all line segments are known to be vectors, there is no good reason for repeating the arrows; they are best dropped altogether. If occasionally we wish to refer to the magnitude

of a vector, it is simpler to use a special symbol; for instance, the vector symbol may be placed between two vertical bars, as in $|OA|$. Simplicity and intelligibility of notation should be the principal concern; too many distinguishing marks often distract the attention from the really important meanings of the symbols.

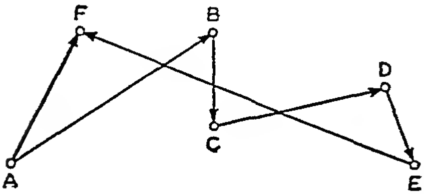

FIG. 7.2. The simplest way to add several vectors is to translate the second vector parallel to itself so that its origin coincides with the end of the first vector, then translate the origin of the third vector to the end of the second, etc. The sum is the vector joining the origin of the first vector to the end of the last.

In this discussion we have implied that the point of application of a vector is not important; we have been considering *free vectors* which, by definition, *are equal if they are parallel, point in the same direction, and are equal in magnitude.* There are also *bound vectors*, attached to their origins. The law of addition of free vectors applies to them only if they have a common origin. There are also *sliding vectors* which are permitted to slide along a straight line without being "changed." Normally the term "vector" is applied to a free vector.

Since the point of application of vectors is irrelevant, vector addition can be performed by taking the end of one vector as the origin of the next. This method is particularly convenient in adding more than two vectors. The sum is the vector drawn from the origin of the first vector to the end of the last; thus, referring to Figure 7.2, we have

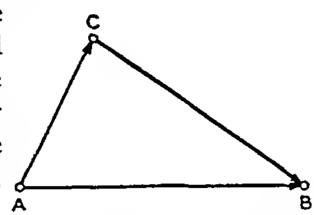

FIG. 7.3. The difference CB of two vectors AB and AC, emerging from a common origin, is the vector joining the end of the second vector to the end of the first.

$$AB + BC + CD + DE + EF = AF. \quad (2)$$

Here the arrows have been dropped and the directions of the vectors are indicated by the order of the letters. If the end of the last vector coincides with the origin of the first, that is, if the vectors form a closed polygon, the sum is zero:

$$AB + BC + CD + DE + EF + FA = 0. \quad (3)$$

A special case of (3) is

$$AB + BA = 0, \qquad BA = -AB; \quad (4)$$

hence we have a definition of subtraction

$$AB - AC = AB + CA = CA + AB = CB. \quad (5)$$

If two vectors are drawn from a common origin, Figure 7.3, the difference

is the vector connecting the end of the second vector with the end of the first.

Although in substance vector addition is different from ordinary addition of numbers, the laws of common algebra, as given in Section 3 of Chapter I, are still satisfied — at least as regards addition and subtraction; multiplication has not yet been defined.

2. *Vector components*

While the sum of any number of vectors is unique, the reciprocal process of breaking up a given vector into its components is indefinite unless further conditions are imposed. Any vector can be uniquely represented as the sum of three vectors in three given directions. This decomposition

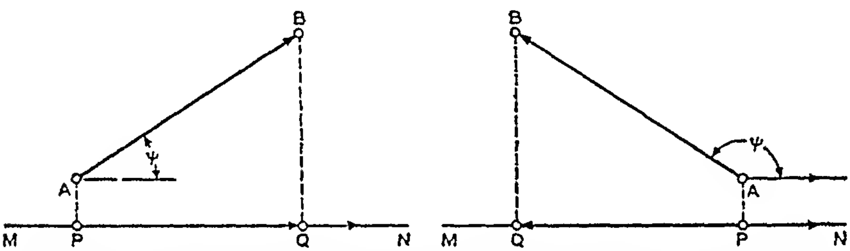

FIG. 7.4. A graphic definition of the component PQ of a vector AB in the direction MN. On the left PQ is positive; on the right it is negative. The component PQ in the opposite direction NM is negative on the left and positive on the right.

is effected by constructing a parallelepiped on the given vector as diagonal, with its faces parallel to the three planes defined by the pairs of given directions. The choice of the three directions is arbitrary (except that they should not be coplanar); but the usual choice is to make them mutually perpendicular.

The *component* of a vector AB along a line MN is the vector PQ connecting the feet of the perpendiculars drawn from the terminals of AB to MN, Figure 7.4. If MN is assigned a positive direction, then PQ is regarded as positive or negative depending on whether it points in the positive or negative direction. The ratio of the length PQ to the magnitude f of AB is the *cosine of the angle* ψ between AB and the positive direction of MN; thus

$$PQ = f \cos \psi. \tag{6}$$

This is really a definition of $\cos \psi$.

If the positive direction of the line MN is chosen from N to M, then the angle between AB and NM is $\pi - \psi$ and

$$PQ = f \cos (\pi - \psi) = -f \cos \psi. \tag{7}$$

The components of a vector AB along the axes of a rectangular frame of reference are called the *direction components* of AB and their ratios to the length ℓ of AB are the *direction cosines* of AB. The direction components are distinguished by subscripts indicating the particular directions in which they are taken. Thus

$$AB_x = x_2 - x_1 = \ell \cos \alpha, \quad AB_y = y_2 - y_1 = \ell \cos \beta,$$
$$AB_z = z_2 - z_1 = \ell \cos \gamma, \quad (8)$$

where α, β and γ are the angles made by AB with the x, y and z axes.

Vectorial components in the *same direction* have the properties of ordinary scalars and *the component of the sum of several vectors is the algebraic sum of the components of the separate vectors.* On the other hand, the components in *different* directions must be added vectorially. In the accepted usage, direction components are thought of primarily as scalars and their vectorial counterparts are designated by means of *unit vectors*. Thus if F_x, F_y and F_z are the direction components of F, we write

$$F = F_x i + F_y j + F_z k, \quad (9)$$

where i, j and k are unit vectors along the respective coordinate axes. This equation implies *multiplication of a vector* by a scalar, meaning multiplication of the magnitude of the vector, which is a scalar, by another scalar, thereby obtaining a second vector of different magnitude but in the same direction as the first.

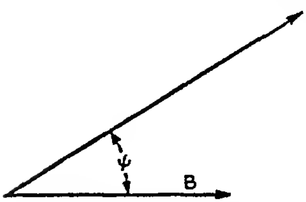

Fig. 7.5. The scalar product of two vectors A, B is defined as the product of their magnitudes and the cosine of the angle ψ.

3. *Scalar product*

A certain function of two vectors is very important in physics. If vector A represents a force acting on a particle, and vector B represents the resulting displacement, Figure 7.5, the work done by A is the product of the magnitudes of the two vectors and the cosine of the angle ψ between them. This product is called the *scalar product* of A and B and is designated by (A,B) or, in Gibbs' notation, by $A \cdot B$. Using the corresponding small letters to denote the magnitudes of the vectors, we write

$$A \cdot B = (A,B) = ab \cos \psi. \quad (10)$$

The direction components of a vector are the scalar products of the vector and the corresponding unit vectors

$$F_x = F \cdot i, \qquad F_y = F \cdot j, \qquad F_z = F \cdot k. \quad (11)$$

It is evident that scalar multiplication is *commutative*. The *distributive*

law of algebraic multiplication is also applicable. This might have been anticipated since the work done by a force is a representative example of the scalar product and the work done by two forces acting on the same particle is equal to the work done by their resultant. On the other hand, the *associative law* of multiplication fails to apply for the simple reason that the scalar product of *three* vectors has no meaning.

The scalar product of two nonvanishing vectors is zero if, and only if, these vectors are perpendicular; thus the condition of perpendicularity of A and B can be expressed as

$$A \cdot B = 0. \tag{12}$$

The scalar product of two *unit* vectors (of unit magnitude), A and B, is the cosine of the angle ψ between them

$$\cos\psi = A \cdot B. \tag{13}$$

For unit vectors along the coordinate axes, we have

$$i \cdot i = j \cdot j = k \cdot k = 1, \qquad i \cdot j = j \cdot k = k \cdot i = 0. \tag{14}$$

By repeated application of the distributive law, we obtain

$$F' \cdot F'' = F'_x F''_x + F'_y F''_y + F'_z F''_z \tag{15}$$

for any two vectors F' and F''. Since the direction components of unit vectors are direction cosines, the cosine of the angle ψ between any two vectors can be expressed in terms of their direction cosines

$$\cos\psi = \cos\alpha' \cos\alpha'' + \cos\beta' \cos\beta'' + \cos\gamma' \cos\gamma''. \tag{16}$$

Rewriting the condition of perpendicularity (12) in a more explicit form, we have

$$A_x B_x + A_y B_y + A_z B_z = 0. \tag{17}$$

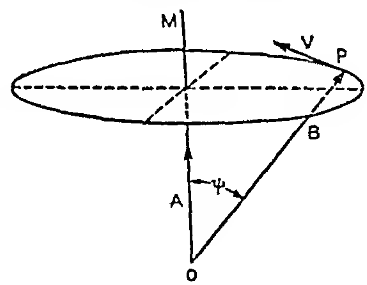

Fig. 7.6. The linear velocity V of a point P in a body rotating about an axis OM is the vector product $A \times B$ where the magnitude of A is the angular speed and B is the vector defining the position of P with reference to some point O on the axis.

4. *Vector product*

Another important function of two vectors is itself a vector. Consider a typical point P of a rigid body rotating with an angular velocity A about OM, Figure 7.6. The linear velocity V of P is a vector perpendicular to A and to any vector B drawn to P from a point on the axis. The speed v of P, that is, the magnitude of the linear velocity, is the product of a, b and the sine of the angle ψ between A and B; thus

$$v = ab \sin\psi. \tag{18}$$

The direction of V is that in which a right-handed screw would advance if turned from A to B. This vector is called the *vector product* of A and B and is denoted by $[A,B]$ or, in Gibbs' notation, by $A \times B$. To give a formal definition: *the vector product of two vectors A and B is a vector perpendicular to both in the direction in which a right-handed screw would advance if turned from A to B; its magnitude in the algebraic sense* is the product of the absolute magnitudes of A and B and the sine of the angle ψ through which A must be turned in order to point in the direction of B.* To rephrase: the vector product of A and B is a vector perpendicular to both, pointing in the direction in which a right-handed screw would advance if turned from A to B *through the smaller* angle between them; its magnitude is the area of the parallelogram constructed on A and B as the sides, Figure 7.7.

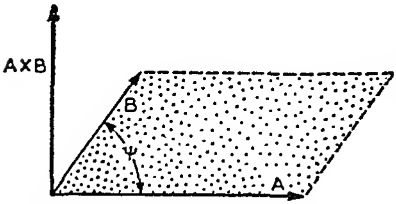

FIG. 7.7. The vector product $A \times B$ is perpendicular to its factors; its magnitude equals the area of the parallelogram with A and B as its sides; its relative direction is as indicated.

The commutative law of multiplication does not apply to vector products but is replaced by a simple modification,

$$A \times B = -B \times A; \tag{19}$$

on the other hand, the distributive law does apply,

$$(A + B) \times C = A \times C + B \times C. \tag{20}$$

For unit vectors along the coordinate axes, we have

$$\begin{aligned} i \times i = j \times j = k \times k = 0, \qquad & i \times j = -j \times i = k, \\ j \times k = -k \times j = i, \qquad & k \times i = -i \times k = j. \end{aligned} \tag{21}$$

Consequently the direction components of the vector product of two vectors F' and F'' are

$$\begin{gathered} (F' \times F'')_x = F'_yF''_z - F'_zF''_y, \qquad (F' \times F'')_y = F'_zF''_x - F'_xF''_z, \\ (F' \times F'')_z = F'_xF''_y - F'_yF''_x. \end{gathered} \tag{22}$$

These formulas include as a special case those for the direction cosines of a vector perpendicular to two given vectors.

Two nonvanishing vectors are parallel if, and only if, their vector product vanishes.

* By magnitude in the algebraic sense is meant the "arithmetic" or "absolute" magnitude with plus or minus sign to distinguish between opposite directions.

Problems

1. Show that $A \cdot (B \times C) = B \cdot (C \times A) = C \cdot (A \times B)$ and that the magnitude of this triple product represents the volume of the parallelepiped whose edges are A, B, C.

2. Prove that $A \times (B \times C) = (A \cdot C)\, B - (A \cdot B)C$. Show directly from geometric considerations that this triple vector product is a vector in the plane defined by B and C.

3. Show that $(A \times B) \cdot (C \times D) = (A \cdot C)(B \cdot D) - (A \cdot D)(B \cdot C)$. *Hint:* Use the results in problems 1 and 2.

4. Prove that $(A \times B) \times (F \times G) = [(A \times B) \cdot G]F - [(A \times B) \cdot F]G$.

5. There is one to one correspondence between two dimensional vectors and complex numbers. Show that the real and imaginary parts of the product $z_1 z_2^*$ are respectively the scalar and vector products of the vectors corresponding to z_1 and z_2.

6. Let the angle between a plane unit vector and the x-axis be φ. By definition the cartesian components of this vector are the circular functions: $\cos\varphi = X$, $\sin\varphi = Y$. Using the formulas for the scalar and vector products of two unit vectors, show that

$$\cos(\varphi_1 - \varphi_2) = \cos\varphi_1 \cos\varphi_2 + \sin\varphi_1 \sin\varphi_2,$$

$$\sin(\varphi_1 - \varphi_2) = \sin\varphi_1 \cos\varphi_2 - \cos\varphi_1 \sin\varphi_2.$$

5. *Invariance*

It is an obvious but nevertheless important fact that scalar and vector products have been so defined that they are independent of the particular coordinate system in which they may have been expressed. They are *invariants* under any transformation of coordinates. Quantities which are not invariant under such transformations possess no intrinsic physical meaning. If the work done by a given force on a particle displaced in a given manner depended on the particular system of coordinates used to specify the force and the displacement, we would not be interested in it.

Definitions such as (10) have the virtue of making the invariance obvious; the magnitudes of vectors and the angle between them are clearly independent of any coordinate system. However, a vector might have been defined as a sort of " complex number " (9); scalar products of complex units might then have been defined by (14); and the scalar product of any two vectors might have been *defined* as subject to the distributive law of multiplication. From these definitions we could derive (15), an equation which possesses but does not exhibit the property of invariance. This property would have to be proved before we could expect the scalar product to be of interest in physical applications.

Any function $f(a,b,\psi)$ of the magnitudes of two vectors A and B and of the angle between them is an invariant; but for the most part such functions would be of no interest in physical applications and would not obey simple algebraic laws. For instance, the " neo-scalar " product, defined as

$ab \tan \psi$, merely for the purpose of illustrating the point, does not obey the distributive law.

6. *Gradient*

Consider a scalar *function of position* $V(x,y,z)$; that is, a function whose values depend on the position of the point under consideration. Temperature, pressure, electric potential are all functions of position. In Sections 9 and 10 of Chapter 5 we considered the rates of change of such functions in different directions. The maximum directional derivative or the *gradient* possesses a direction as well as a magnitude; since it is the limit of $\Delta V/\Delta n$, where Δn is the normal distance between the level surfaces, the gradient is independent of the frame of reference; and to prove that it is really a vector we have only to show that the addition of gradients follows the rules of vectorial addition.

Thus let $V(x,y,z) = V_1(x,y,z) + V_2(x,y,z)$ be the sum of two functions of position. In Section 2 we have seen that the component (in any given direction) of the sum of two vectors is equal to the sum of their components and that any vector is the sum of its components in three mutually perpendicular directions. Equations (5–30) and (5–31) show that the partial derivatives are the components of the gradient along the coordinate axes. These components are certainly added algebraically since

$$\frac{\partial V}{\partial x} = \frac{\partial(V_1 + V_2)}{\partial x} = \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial x}; \tag{23}$$

and similarly for the remaining components; thus grad V is indeed a vector.

Willard Gibbs introduced the following vector operator " del "

$$\nabla = i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z} \tag{24}$$

and used it to denote the gradient

$$\text{grad } V = \nabla V = \left(i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}\right)V = i\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y} + k\frac{\partial V}{\partial z}. \tag{25}$$

This notation is tied to cartesian coordinates and is not very useful in physical thinking; but it is a valuable tool in the rapid derivation of some important formulas. As a general policy a single standard notation is to be highly recommended; but it is just as important to admit exceptions when their usefulness has been proved.

Problem. Find directly from the definition grad (r^n), where r is the distance from a fixed point.

Ans. The gradient is a radial vector of magnitude nr^{n-1}.

7. Divergence

Associating a vector with every point in a given region of space, we obtain a *vector point function* or a *vector field*. Electric intensity, magnetic intensity, fluid velocity are all examples of vector point functions. The concept of divergence may be illustrated by visualizing the vector field $F(x,y,z)$ as fluid flow; at each point the vector F is assumed to represent the quantity of fluid passing per unit area perpendicular to the lines of flow (in some fixed interval of time). The flux Φ of F across a given surface S, that is, the quantity of fluid passing through S, is the surface integral

$$\Phi = \int\int F_n \, dS = \int\int |F| \cos (F,n) \, dS, \tag{26}$$

where F_n is the normal component of F (the component of F along the normal to dS) and (F,n) is the angle formed by F and the normal to dS. If n represents the unit normal, then $F_n = F \cdot n$. The element of area itself can be regarded as a vector $\overrightarrow{dS}$ directed along the normal, with dS representing its magnitude. These considerations lead to two additional symbolic forms of (26):

$$\Phi = \int\int F \cdot n \, dS = \int\int F \cdot \overrightarrow{dS}. \tag{27}$$

No matter how the integrand is written, it is the product of the magnitude of the vector, the elementary area in the ordinary sense, and the cosine of the angle between the vector and the normal.

The outward flux of F through a simply connected closed surface S divided by the enclosed volume v is the *average divergence* of F throughout the volume

$$\text{av div } F = \frac{\int\int F_n \, dS}{v}. \tag{28}$$

The limit of the average divergence, as S shrinks to a point, is the *divergence of F at the point*

$$\text{div } F = \lim \frac{\int\int F_n \, dS}{v}. \tag{29}$$

If we divide the total volume v enclosed by the surface S into elementary volumes, we observe that the total flux of F across S is the sum of the fluxes through the boundaries of the elementary volumes, the fluxes

through the common partitions between the elements contributing nothing to the whole; thus

$$\iint F_n\,dS = \iiint \operatorname{div} F\,dv. \tag{30}$$

In simple cases the divergence may best be calculated directly from the definition; otherwise the definition is applied first to obtain expressions for the divergence in terms of the partial derivatives of the vector components. Let F be a field of vectors directed along the radii emerging from a fixed point O and suppose that the magnitude is equal to r, the distance from O. It is clear that div F will depend only on r and will be the same at any point on a sphere of radius r. Thus instead of applying our definition to a small sphere surrounding the point at which div F is to be determined we may consider a vanishingly thin spherical shell enclosing the point between two spheres of radii r and $r + \Delta r$. The flux across the sphere of radius r is $(4\pi r^2)r = 4\pi r^3$; the flux across the sphere of radius $r + \Delta r$ is $4\pi(r + \Delta r)^3$; the flux out of the spherical shell is $12\pi r^2\Delta r +$ small quantities of higher order; the volume of the shell is $4\pi r^2\Delta r +$ small quantities of higher order; the flux/volume ratio is $3 +$ small quantities of higher order; hence div $F = 3$, everywhere except perhaps at the origin where the above method is not applicable. But at this point we can apply the definition to a small sphere surrounding the origin and we find that the flux/volume ratio is $(4\pi r^2)r/(4/3)\pi r^3 = 3$, and div $F = 3$ everywhere.

Suppose now that the magnitude of F is $1/r^2$. Following the above procedure we obtain div $F = 0$ everywhere *except* at the point $r = 0$. But at the origin the outward flux through a small sphere of radius r is $(4\pi r^2)(1/r^2) = 4\pi$; the volume of the sphere is $(4/3)\pi r^3$; and the ratio becomes infinite as r approaches zero. Thus at $r = 0$ div $F = \infty$.

In deriving equation (30) we have naturally assumed that we could carry out the instructions for calculating both sides. If F is defined as in the above example and the region includes the point $r = 0$, we shall not be able to carry out the instructions on the right-hand side. In fact, in order to obtain (30) we have to argue as follows:

1. The flux across the boundary of the entire volume is the sum of the fluxes across the boundaries of the elementary volumes,
2. The contributions from boundaries *common* to elementary volumes are nil,
3. The limit (29) exists in the interior of the volume.

The second part of the argument requires F to be single-valued and the third part states explicitly the condition for the existence of div F. If we think of div $F\,dv$ not as a product of two quantities but as a single quantity,

the third part of the argument is not needed and (30) will be more generally true.

The case of two-dimensional vector fields can be treated similarly. The *average surface divergence* is defined as the outward flux of F across the boundary s of an area S divided by S and the limit of this quotient as s contracts to a point is defined as the *surface divergence at the point*

$$\text{sur div } F = \lim \frac{\int F_n \, ds}{S}. \tag{31}$$

The prefix " sur " need be used only on these rare occasions when three- and two-dimensional types of divergence occur together. The formula corresponding to (30) is

$$\int F_n \, ds = \int\int \text{sur div } F \, dS. \tag{32}$$

The *linear divergence* is an ordinary derivative and the formula corresponding to (30) and (32) is

$$F(b) - F(a) = \int_a^b \frac{dF}{dx} \, dx. \tag{33}$$

The electric current in a wire is the time rate of flow of electric charge past a given point and the rate at which the charge per unit length is accumulating is the negative of the linear divergence of the current.

Each time the dimensionality of the vector field is decreased by one, the dimensionalities of the integrals in these formulas are also decreased by one.

8. *Calculation of divergence in cartesian coordinates*

If the vector field is expressed in cartesian coordinates, then to evaluate its divergence we may choose an elementary parallelepiped formed by coordinate planes, Figure 7.8, and let it shrink to the center $P(x,y,z)$. Let F_x, F_y, F_z be the components of the vector at P and Δx, Δy, Δz be the edges of the parallelepiped. The flux across the cross section $\Delta y \Delta z$ at P is $F_x \Delta y \Delta z$ + small quantities of higher order; hence the flux out of the cube across the front and back faces is $D_x(F_x \Delta y \Delta z)\Delta x$. Since $\Delta y \Delta z$ remains

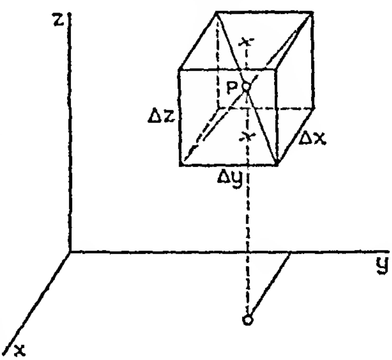

FIG. 7.8. An elementary coordinate cell for calculating the divergence of a vector in cartesian coordinates.

constant as we slide the section from back to front, the flux is $D_xF_x\Delta x\Delta y\Delta z$. Similar expressions are obtained for the fluxes through the left and right faces and through the top and bottom. Dividing the sum of these fluxes by the volume $\Delta x\Delta y\Delta z$ of the parallelepiped we have

$$\operatorname{div} F = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}. \tag{34}$$

In cartesian coordinates (30) assumes the following forms

$$\begin{aligned}\iint [F_x \cos(x,n) + F_y \cos(y,n) + F_z \cos(z,n)]\, dS \\ = \iiint \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\right) dx\, dy\, dz, \\ \iint (F_x\, dy\, dz + F_y\, dx\, dz + F_z\, dx\, dy) \\ = \iiint \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\right) dx\, dy\, dz.\end{aligned} \tag{35}$$

In the second form the element of area $dy\,dz$ should be counted as positive or negative depending on whether the angle (x,n) between the normal and the x-axis is acute or obtuse; similarly for the other elements of area. The above equations constitute *Green's theorem.* Any set $P(x,y,z)$, $Q(x,y,z)$, $R(x,y,z)$ of three scalar functions of position can be chosen as components of a vector and the above equations represent general relationships between certain volume and surface integrals.

In Section 6 the vector operator *del* was introduced. If we treat it as a vector and multiply it scalarly by a vector F, we find

$$\begin{aligned}\nabla \cdot F &= \left(i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}\right)\cdot(iF_x + jF_y + kF_z) \\ &= \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = \operatorname{div} F.\end{aligned} \tag{36}$$

For this reason $\nabla\,\cdot$ may be used as an alternate for div; however, the notation is tied too specifically to cartesian coordinates.

Problems

1. Let F be a radial vector whose magnitude is $1/r^2 = 1/(x^2 + y^2 + z^2)$. Find its cartesian components and use (34) to obtain div F. Note the complications which arise when the frame of reference is not adapted to the simplicity of the vector field. What happens at $r = 0$?

2. Let F be a radial vector whose magnitude is r^n. Show that div $F = (n+2)r^{n-1}$ by direct application of (29) and by (34).

3. Let F be directed along lines perpendicular to a fixed line and $|F| = \rho^n$, where ρ is the distance from the axis. Show that div $F = (n+1)\rho^{n-1}$ by the direct and indirect methods. Discuss the case $n = -1$.

4. Show that

$$\begin{aligned} \int [P\cos(x,n) + Q\sin(x,n)]\,ds &= \int\int\left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)dx\,dy, \\ \oint (P\,dy - Q\,dx) &= \int\int\left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)dx\,dy. \end{aligned} \tag{37}$$

9. *Curl*

The *line integral of a vector F along a curve AB*, Figure 7.9, *is defined as the line integral* $\int_{AB} F_s\,ds$, *where F_s is the tangential component of F and ds is the differential element of length.* Evidently, $\int_{AB} F_s\,ds = \int_{AB} F \cdot \overrightarrow{ds}$, if we regard the element of length along the curve as a vector tangential to the curve. If F is a force, the line integral represents the work done by F on a particle moving along AB. If the direction of integration is reversed, the line integral changes its sign

$$\int_{AB} F_s\,ds = -\int_{BA} F_s\,ds. \tag{38}$$

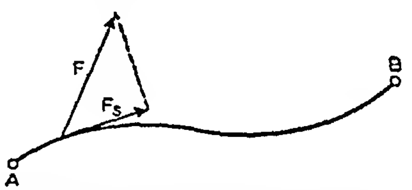

Fig. 7.9. The line integral of a vector F along AB is defined as the line integral of the tangential component F_s along the curve.

When the curve is closed the positive value of the line integral is called the *circulation* of F and the direction of integration corresponding to this value the *direction of circulation*. The circulation of the electric intensity round a given curve represents the total voltage which would be impressed on a conducting wire coinciding with the curve. The circulation of any force represents the work done on a particle moving round a closed curve. If the circulation vanishes for every closed curve, the field of force is *conservative;* the field does no net work on a particle returning to its original position.

The circulation round an infinitely small plane loop depends upon its size and orientation. As the loop contracts to a point while keeping its orientation unaltered, the circulation per unit area may approach a limit. This limit is called the *component of the curl of F* in that direction, normal to the plane of the loop, in which the circulation appears to be clockwise.

For a certain orientation of the loop the circulation per unit area is maximum; this maximum value combined with the direction in which the circulation appears to be clockwise is denoted by curl F (read " curl of F"). Presently we shall show that curl F is a vector.

If F is a field of force, the circulation is the work done by the field when a particle is carried round a closed path and curl F represents the maximum work per unit area. The magnetic intensity H is defined as the force which would be exerted on a unit magnetic pole and the magnetomotive force U is the work which would be performed by H when a unit pole is carried along a given curve. According to Ampère the magnetomotive force round a closed curve is equal to the total electric current passing across any surface bounded by this curve; hence the curl of the magnetic intensity is equal to the density J of the electric current (that is, the current per unit area normal to the lines of flow). Symbolically

$$\text{curl}\, H = J. \tag{39}$$

Similarly, Faraday's law of electromagnetic induction may be stated in the form

$$\text{curl}\, E = -\frac{\partial B}{\partial t}, \tag{40}$$

where E is the electric intensity, B is the density of the magnetic flux, and t is time. The left side of the equation is the electromotive force per unit area of a closed loop, or the work done by E in carrying a unit electric charge round the loop divided by the area of the loop; the right side is the negative of the time rate of change of the magnetic flux through the loop divided by the area of the loop. The difference in sign in (39) and (40) expresses the fact that if J and $\partial B/\partial t$ are similarly directed, then, looking in this direction, the magnetomotive and electromotive forces appear respectively clockwise and counterclockwise.

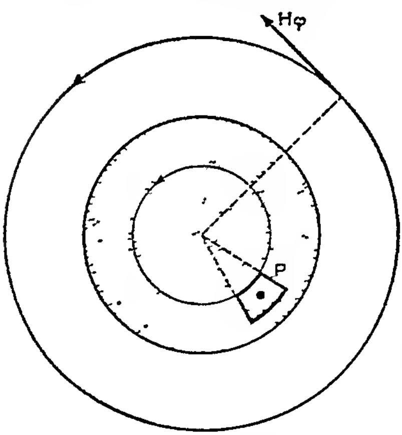

FIG. 7.10. A cross section of a cylindrical conductor carrying an electric current (the shaded area), two typical magnetic lines of force (thin circles), and an element of area round a point P.

Consider a cylindrical wire carrying a uniformly distributed electric current of density J; Figure 7.10 shows the cross section of the wire and the circular lines of magnetic intensity (they are counterclockwise if the current is toward the reader). Inside the wire the current within a circle of radius ρ is $\pi\rho^2 J$; according to Ampère this should equal the magnetomotive force $2\pi\rho H_\varphi$ (symmetry requires the independence of H_φ from the angular

coordinate); therefore, $H_\varphi = \frac{1}{2}\rho J$. For any ρ greater than the radius a of the wire, the current enclosed by the magnetic line is $\pi a^2 J$ and $H_\varphi = a^2 J/2\rho$. To summarize

$$H_\varphi = \tfrac{1}{2}\rho J, \quad \rho \leq a; \\ = \frac{a^2 J}{2\rho}, \quad \rho \geq a. \tag{41}$$

The radial component H_ρ and the axial component H_z of the magnetic field of the straight current filament are both zero.

Let us now verify that (39) is true for all points P, Figure 7.10. Choosing an elementary area bounded by two radii and two circles, we find that the total counterclockwise magnetomotive force round the boundary is $(H_\varphi + \Delta H_\varphi)(\rho + \Delta\rho)\Delta\varphi - H_\varphi\rho\Delta\varphi = \rho(\Delta H_\varphi)\Delta\varphi + H_\varphi\Delta\rho\Delta\varphi$ plus an infinitesimal of higher order. The area is $\rho\Delta\rho\Delta\varphi$ and the magnetomotive force per unit area is $\Delta H_\varphi/\Delta\rho + H_\varphi/\rho$ and in the limit $dH_\varphi/d\rho + H_\varphi/\rho$. By (41), the sum of these terms is J for any point inside the wire; outside the wire the two terms cancel each other. Let us note carefully that the magnetic lines are circular both inside and outside the current-bearing wire and that the total circulation round these lines does not vanish anywhere outside the axis $\rho = 0$; but the curl of the vector field is different from zero inside the wire and vanishes outside. The curl expresses the circulation in the vicinity of a point. An example from mechanics will illustrate this feature.

Consider the rotation of a solid body about an axis, Figure 7.6. The field of the linear velocity is

$$v_\rho = 0, \qquad v_\varphi = \omega\rho, \qquad v_z = 0, \tag{42}$$

where ω is the angular velocity and ρ the distance from the axis. This field is similar to (41) with $a = \infty$; hence, if we regard the angular velocity as a vector along the axis of rotation,

$$\vec{\omega} = \tfrac{1}{2} \operatorname{curl} \vec{v}. \tag{43}$$

In this case, not only is the whole body revolving about OM with the angular velocity ω but each element is revolving with this velocity about an axis passing through the center of the element.

Returning to the definition of curl F let us calculate its components in cartesian coordinates. To obtain the z-component, $\operatorname{curl}_z F$, we take an elementary rectangle in the xy-plane, with its sides formed by coordinate lines, Figure 7.11, and calculate the counterclockwise circulation round the boundary. The line integral of F along CD is $F_y\Delta y$ + infinitesimals of higher order, where F_y is taken at the center of the rectangle; the rate of change of this quantity in the x-direction is $D_x(F_y\Delta y)$, omitting the higher

order terms which will disappear in the final calculation, and the difference between the line integrals along LN and KM is $D_x(F_y\Delta y)\Delta x$. This difference is the contribution to the counterclockwise circulation from LN and MK. Similarly, $D_y(F_x\Delta x)\Delta y$ is the contribution to the *clockwise* circulation from MN and LK. The total counterclockwise circulation is thus obtained and the limit of its ratio to the area $\Delta x\Delta y$ is

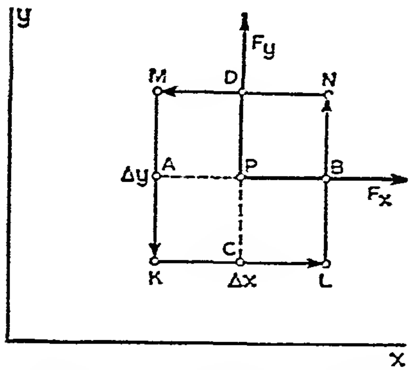

FIG. 7.11. An elementary rectangle bounded by coordinate lines for calculating the curl of a vector at a point P.

$$\operatorname{curl}_z F = \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}. \tag{44}$$

In the partial derivative $D_x(F_y\Delta y)$ we note that Δy is constant as we slide CD from one side of the rectangle to the other; hence Δy may be moved outside the derivative sign. In curvilinear coordinates the element of length will vary from one side of the corresponding curvilinear rectangle to the other and certain factors depending on the position of the point will appear in the vector components (see Chapter 8).

Permuting x, y, z cyclically we have

$$\begin{aligned} \operatorname{curl}_x F &= \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}, \\ \operatorname{curl}_y F &= \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}. \end{aligned} \tag{45}$$

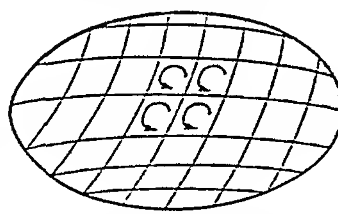

FIG. 7.12. The circulation of a vector round the periphery of a surface is equal to the sum of the circulations round the peripheries of the elements of the surface.

It is evident that the components of the curl of the sum of two vectors are the sums of the corresponding components of the curls of the individual vectors and that curl F is a vector.

Consider a surface S bounded by a *simple* closed curve (a curve which does not intersect itself). If S is divided into elementary areas, Figure 7.12, the circulation round the boundary s of S is the sum of the circulations round the boundaries of the elementary areas, since the contributions from the boundaries common to adjacent elements cancel out (assuming that the vector function is single valued); hence

$$\int F_s\, ds = \int\int \operatorname{curl}_n F\, dS, \tag{46}$$

where the integration on the left appears clockwise to an observer looking

through S in the direction of the normal. If $\overrightarrow{dS}$ is a vector normal to the surface and $\overrightarrow{ds}$ tangential to the boundary, then

$$\int F \cdot \overrightarrow{ds} = \int\int (\text{curl } F) \cdot \overrightarrow{dS}. \tag{47}$$

This is Stokes' theorem. In cartesian coordinates

$$\begin{aligned}\int (F_x\,dx + F_y\,dy + F_z\,dz) = \int\int \Bigg[&\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) \cos(n,x) \\ &+ \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) \cos(n,y) + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) \cos(n,z)\Bigg]\,dS.\end{aligned} \tag{48}$$

Any set of three functions P, Q, R may be regarded as the components of a vector F, and we have a theorem about the equivalence of certain line and surface integrals.

If $F_x = P(x,y)$, $F_y = Q(x,y)$, $F_z = 0$, then

$$\oint (P\,dx + Q\,dy) = \int\int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy. \tag{49}$$

This is essentially the same as equation (37).

We have seen that div F is the scalar product of *del* and F; the student can easily verify that curl F is the vector product

$$\text{curl } F = \nabla \times F. \tag{50}$$

10. *Some vector identities*

There are numerous identities connecting the gradient, divergence, curl, and *laplacian*

$$\Delta V = \text{div grad } V = \nabla \cdot \nabla V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \tag{51}$$

The following list provides a few examples:

$$\text{div curl } F = 0, \qquad \text{curl grad } V = 0; \tag{52}$$

$$\text{curl curl } F = \text{grad div } F - \Delta F; \tag{53}$$

$$\text{div } (VF) = V \text{ div } F + F \cdot \text{grad } V; \tag{54}$$

$$\text{curl } (VF) = V \text{ curl } F - F \times \text{grad } V; \tag{55}$$

$$\text{div } (F \times G) = G \cdot \text{curl } F - F \cdot \text{curl } G; \tag{56}$$

$$\text{curl } (F \times G) = F \text{ div } G - G \text{ div } F + (G \cdot \nabla)F - (F \cdot \nabla)G. \tag{57}$$

It is evident from the context that F and G are vectors while V is a scalar. The laplacian of a vector F is defined as a vector whose components are the laplacians of the components of F.

The above identities can be verified in cartesian coordinates; for example,

$$\operatorname{div} \operatorname{curl} F = \frac{\partial}{\partial x}\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) = 0.$$

In some cases Gibbs' notation is suggestive. Thus the second part of (52) is $\nabla \times \nabla V = 0$; the truth of this equation is strongly suggested since the vector product of two equal vectors vanishes.

That div curl F vanishes can be deduced directly from (30) and (46). Replace F by curl F in the former equation,

$$\iiint \operatorname{div} \operatorname{curl} F\, dv = \iint \operatorname{curl}_n F\, dS; \tag{58}$$

exclude a small area of the surface of integration and apply (46) to the remainder of the surface; the surface integral is equal to the line integral round the edge of the excluded area and this approaches zero as the area shrinks to a point.

Next consider (54)

$$\begin{aligned} \operatorname{div} (VF) &= \frac{\partial VF_x}{\partial x} + \frac{\partial VF_y}{\partial y} + \frac{\partial VF_z}{\partial z} \\ &= V\frac{\partial F_x}{\partial x} + F_x\frac{\partial V}{\partial x} + V\frac{\partial F_y}{\partial y} + F_y\frac{\partial V}{\partial y} + V\frac{\partial F_z}{\partial z} + F_z\frac{\partial V}{\partial z}. \end{aligned}$$

The sum of the first, third and fifth terms represents the first term on the right of (54); the remaining terms represent the second term.

11. *Green's theorems*

A number of useful modifications of Green's formula (30) can be obtained by various substitutions. Let $F = U \operatorname{grad} V$, where U and V are two scalar functions; then

$$\iiint \operatorname{div} (U \operatorname{grad} V)\, dv = \iint U \operatorname{grad}_n V\, dS. \tag{59}$$

From (54) we obtain

$$\begin{aligned} \operatorname{div} (U \operatorname{grad} V) &= U \operatorname{div} \operatorname{grad} V + (\operatorname{grad} V) \cdot (\operatorname{grad} U) \\ &= U\Delta V + (\operatorname{grad} U) \cdot (\operatorname{grad} V). \end{aligned} \tag{60}$$

Substituting from (60) in (59)

$$\iiint U\Delta V\, dv = \iint U \operatorname{grad}_n V\, dS - \iiint (\operatorname{grad} U) \cdot (\operatorname{grad} V)\, dv. \tag{61}$$

Interchanging U and V and subtracting the result from (61), we obtain a symmetrical form

$$\iiint (U\Delta V - V\Delta U)\, dv = \iint \left(U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n}\right) dS. \tag{62}$$

12. *Irrotational, solenoidal and general fields*

If the circulation of a vector field vanishes for *every* closed curve, the field is called *irrotational* or *lamellar*. Gravitational and electrostatic fields are examples. Such fields can be expressed as gradients of single-valued *potential functions*

$$F = \text{grad } V. \tag{63}$$

This is the form employed in gravitational theory; in electrostatics the usual form is $F = -\text{grad } V$.

If the flux of a vector field vanishes for *every* closed surface, the field is called *solenoidal*. The magnetic field is an example. Such fields may be expressed as curls of single-valued *vector potentials*

$$F = \text{curl } A. \tag{64}$$

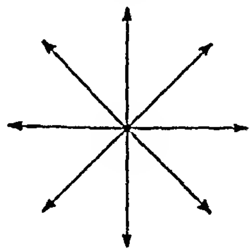

FIG. 7.13. An irrotational field.

A general field may be expressed as the sum of irrotational and solenoidal fields

$$F = \text{grad } V + \text{curl } A. \tag{65}$$

The word "every" in the above definitions is essential. Consider the field of a positively charged particle, Figure 7.13. This field is radial and it varies inversely as the square of the distance r from the particle

$$F_r = \frac{k}{r^2}; \tag{66}$$

it is the gradient of the following potential

$$V = -\frac{k}{r}. \tag{67}$$

The field is irrotational because the circulation round every curve is zero; this is easily verified for curves bounded by the lines of force and circles concentric with the particle. Green's theorem extends the conclusion to all closed curves.

A very similar field is the magnetic field of electric currents circulating around the surface of a semi-infinite cylinder of infinitely small radius, Figure 7.14 The magnetic flux comes out of the end of the cylinder and spreads

out radially. The appearance of the field is similar to that in Figure 7.13 everywhere except along the radius occupied by the cylinder where the flux is toward the center of the field. It can be shown that the field (66) is expressible by (64) if

$$A_r = 0, \qquad A_\varphi = -\frac{k \cot \theta}{r}, \qquad A_\theta = 0, \tag{68}$$

where the only nonvanishing component is along the circles coaxial with the cylinder. This example shows that the same field is sometimes expressible almost everywhere either as the gradient of a scalar function or as the curl of a vector function; it is "almost" that may make the difference between the field being irrotational or solenoidal. It is possible for a field to be both irrotational and solenoidal; an example is $F =$ constant.

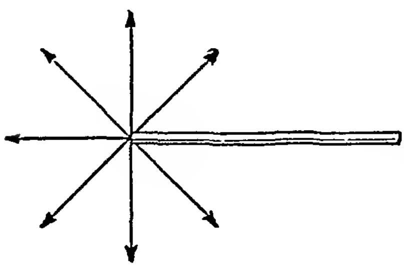

FIG. 7.14. A solenoidal field similar to the irrotational field in Figure 7.13.

If the circulation of F vanishes round every curve, the line integral of F depends only on the end points A, B of the path of integration; any two paths joining A and B form a closed curve and the circulation round this curve is the difference between the line integrals. Keeping A fixed, we define a scalar single-valued function which depends only on the position of B

$$V = \int_{AB} F_s \, ds. \tag{69}$$

The differential element of this function and the directional derivative are

$$d_s V = F_s \, ds, \qquad F_s = \frac{d_s V}{ds} = \frac{\partial V}{\partial s}; \tag{70}$$

that is, $F = \operatorname{grad} V$.

The fields of continuous distributions of mass or electric charge may be regarded as resultants of the fields of volume elements of mass or charge; but while direct addition of the fields is complicated by the necessity of taking their directions into consideration, the addition of the potentials is a simple integration over the volume occupied by the "sources" (mass or charge). The potential of each element is given by (67) where k is proportional to the mass or charge of the element. The latter is proportional to the volume of the element; hence,

$$V = \iiint \frac{k' \, dv}{r}, \tag{71}$$

where k' is proportional to the density of mass or charge. The field can now be found by differentiating V.

On the other hand, taking the divergence of (63), we have *Poisson's differential equation* for V

$$\Delta V = \operatorname{div} F, \tag{72}$$

and the potential should be expressible in terms of div F. By analogy it seems possible that we might be able to find a solution of (72) similar in form to (71). Thus, turning to equation (62), we observe that the left side will be an integral depending on div F if U satisfies *Laplace's equation*

$$\Delta U = 0. \tag{73}$$

The function $U = 1/r$ satisfies this equation everywhere except at $r = 0$ where neither U nor its derivatives exist. Thus $\Delta U = \operatorname{div} \operatorname{grad} U$ and grad $(1/r)$ is a radial vector, $-1/r^2$, the divergence of which vanishes everywhere except at $r = 0$. Hence (62) can be applied to any volume which excludes $r = 0$. The stage is now set for the final steps. Let

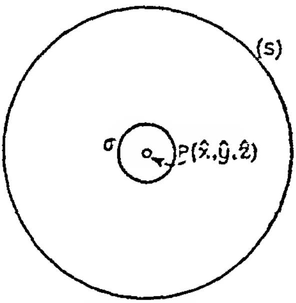

FIG. 7.15. Illustrating the proof of equation (76).

$$U = \frac{1}{r} = \frac{1}{\sqrt{(x - \hat{x})^2 + (y - \hat{y})^2 + (z - \hat{z})^2}}, \tag{74}$$

where $(\hat{x},\hat{y},\hat{z})$ is some fixed point in space; U is a solution of (73) everywhere except at this point; we exclude this point from the volume of integration by a small sphere σ as shown in Figure 7.15; and we apply (62) to the volume bounded by σ and a closed surface S which is to be expanded to infinity and which is also chosen to be a sphere for convenience of integration

$$\iiint \frac{\Delta V}{r}\, dv = \iint\limits_{S+\sigma} \left(\frac{1}{r}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\frac{1}{r}\right) dS. \tag{75}$$

As S recedes to infinity, the integral on the right, taken over S, will approach zero if at infinity V varies as r^{-m} where $m > 0$. If grad $V = F$ and is finite, the first part of the surface integral vanishes over σ as the radius approaches zero; this is true since the surface of the sphere is $4\pi r^2$. On σ, $D_n(1/r) = -D_r(1/r) = 1/r^2$, the negative sign being due to the fact that in (62) n is the normal pointing out of the volume; hence the contribution of the second part of the integral is $-4\pi V(\hat{x},\hat{y},\hat{z})$ and

$$V(\hat{x},\hat{y},\hat{z}) = -\frac{1}{4\pi}\iiint \frac{\Delta V}{r}\, dv = -\frac{1}{4\pi}\iiint \frac{\operatorname{div} F}{r}\, dv. \tag{76}$$

It is not very difficult to extend this result and express a general class of vector fields in the form (65). Taking the divergence and curl of (65), we have

$$\operatorname{div} F = \Delta V, \qquad \operatorname{curl} F = \operatorname{curl} \operatorname{curl} A = \operatorname{grad} \operatorname{div} A - \Delta A. \tag{77}$$

The field F depends on three functions (its cartesian components) but on the right of (65) we have four functions; one of the latter is redundant and we are free to restrict A by making $\operatorname{div} A = 0$. Thus the above equations are reduced to

$$\Delta V = \operatorname{div} F, \qquad \Delta A = -\operatorname{curl} F. \tag{78}$$

Making use of (76), we obtain

$$F = -\frac{1}{4\pi} \operatorname{grad} \iiint \frac{\operatorname{div} F}{r}\, d\sigma + \frac{1}{4\pi} \operatorname{curl} \iiint \frac{\operatorname{curl} F}{r}\, d\sigma. \tag{79}$$

Naturally, (79) is true only if the volume integrals converge; but the resolution (65) is possible even if they do not. In the latter case we have to look for another method of obtaining V and A.

It will be recalled that V was assumed to vanish at infinity in a certain way; the assumption was made in order to eliminate the surface integral over S by extending the integration to infinity. If S is kept finite and the corresponding surface integral is retained in (75), a more general formula is obtained. In this formula, however, V and A are not determined solely by $\operatorname{div} F$ and $\operatorname{curl} F$ but by the values of F and its derivatives over S.

Furthermore, we have assumed that F and its derivatives are finite. Even such simple fields as that of a point source are not included in (79) for at the source F is infinite. Equation (79) represents the field of a distribution of sources when the volume density is finite; to this we should add surface integrals expressing the field of surface distributions of sources, if any; line integrals of line distributions of sources; and finally, a sum representing the field of a system of isolated point sources. The surface integral over S expresses the field of sources lying outside S. When all these terms are included, a perfectly general expression is obtained for the field in the *interior* of S.

Of course, we could limit the formula to volume integrals on the ground that in nature all volume densities are finite; this would result however in unnecessary complications. Small particles, thin wires and thin sheets actually exist; and frequently nothing is gained by bringing in the small dimensions.

In special cases (79) takes on more familiar forms. If F is the electric intensity E, $\operatorname{div} E$ equals the density of charge q divided by the dielectric

constant ϵ; curl E vanishes and (79) reduces to

$$E = -\text{grad} \int\int\int \frac{q}{4\pi\epsilon r} dv. \tag{80}$$

The integral is called the *electric potential.* On a conductor the charge is concentrated on the surface and the integral becomes a surface integral.

If F is the magnetic intensity H of the field generated by electric currents, div $H = 0$ and in view of (39) we have

$$H = \text{curl} \int\int\int \frac{J}{4\pi r} dv. \tag{81}$$

The integral is called the *magnetic vector potential.*

the meridian plane passing through P, and the distance z from the equatorial plane.

In the spherical system, $P(r,\theta,\varphi)$ is specified by the distance r from the origin, the *polar angle* θ made with the polar axis by the radius OP, and the longitude φ.

Some current notations for cylindrical and spherical coordinates are mutually conflicting in that the same letter r is used to denote the distance from the polar axis in one system and the distance from the origin in the other; likewise, the letter θ is sometimes used to denote the polar angle in spherical coordinates and the longitude in cylindrical coordinates. Such conflicting symbolism should be avoided since there are problems, increasing in number, in which it is convenient to employ both systems simultaneously.

The following are the equations for the transformation of coordinates in the three systems:

$$
\begin{aligned}
&x = \rho \cos\varphi = r \sin\theta \cos\varphi; && \rho = \sqrt{x^2 + y^2} = r \sin\theta; \\
&y = \rho \sin\varphi = r \sin\theta \sin\varphi; && \varphi = \tan^{-1}(y/x) = \varphi; \\
&z = z = r \cos\theta; && z = z = r\cos\theta; \qquad (1) \\
&r = \sqrt{x^2 + y^2 + z^2} = \sqrt{\rho^2 + z^2}; && \\
&\theta = \tan^{-1}(\sqrt{x^2 + y^2}/z) = \tan^{-1}(\rho/z); && \\
&\varphi = \tan^{-1}(y/x) = \varphi. &&
\end{aligned}
$$

Problems

1. Find the cartesian components of a unit vector in the direction (θ,φ).

Ans. $X = \sin\theta\cos\varphi, \quad Y = \sin\theta\sin\varphi, \quad Z = \cos\theta.$

2. Using the formula for the scalar product of two unit vectors in the directions (θ_1,φ_1) and (θ_2,φ_2), show that the cosine of the angle ψ between these directions is

$$\cos\psi = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2).$$

2. *Differential elements of length, area, volume*

In the cartesian system the coordinates represent distances along coordinate lines, measured from points of intersection with the principal reference planes; hence the distances between points on the same coordinate line, such as $P(x_1,y,z)$ and $P(x_2,y,z)$, are simply the differences between the corresponding coordinates

$$s_x = x_2 - x_1; \qquad s_y = y_2 - y_1; \qquad s_z = z_2 - z_1. \tag{2}$$

The distance between any two points $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ is given by the theorem of Pythagoras

$$s = \sqrt{s_x^2 + s_y^2 + s_z^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \quad (3)$$

On the differential scale these formulas become

$$ds_x = dx, \qquad ds_y = dy, \qquad ds_z = dz;$$
$$ds = \sqrt{ds_x^2 + ds_y^2 + ds_z^2} = \sqrt{dx^2 + dy^2 + dz^2}. \quad (4)$$

These simple formulas hold only if all distances are measured in the same units. If x is measured in feet, y in inches and z, s in centimeters, then in centimeter units equations (4) become

$$ds_x \simeq 30dx, \qquad ds_y \simeq 2.5dy, \qquad ds_z = dz;$$
$$ds \simeq \sqrt{900dx^2 + 6.25dy^2 + dz^2}. \quad (5)$$

On doubly logarithmic coordinate paper the distances along the coordinate lines are proportional to the logarithms of the coordinates and in the expressions for the differential distances in terms of the differentials of the coordinates the scale factor varies from point to point. A similar situation arises in any curvilinear system of coordinates, Figure 8.2; the differential distances along coordinate lines are proportional to the differentials of coordinates

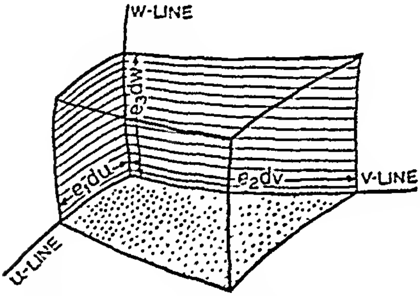

FIG. 8.2. An elementary coordinate cell formed by the intersection of neighboring coordinate surfaces. The edges of the cell are proportional to the differentials of the corresponding coordinates.

$$ds_u = e_1\,du,$$
$$ds_v = e_2\,dv, \quad (6)$$
$$ds_w = e_3\,dw;$$

the coefficients e_1, e_2, e_3 depend on the position of the point in whose vicinity the distances are being evaluated. If the system is *orthogonal*, the distance between two infinitely close points is

$$ds = \sqrt{ds_u^2 + ds_v^2 + ds_w^2} = \sqrt{e_1^2\,du^2 + e_2^2\,dv^2 + e_3^2\,dw^2}, \quad (7)$$

since the curvilinear triangles approach the rectilinear as their size decreases; but there is no simple formula for the distance between distant points.

A general expression for the differential element of length can be obtained as follows. By definition, u, v, w are three families of surfaces:

$$u = f(x,y,z), \qquad v = g(x,y,z), \qquad w = h(x,y,z). \quad (8)$$

Solving for x, y, z we have

$$x = F(u,v,w), \qquad y = G(u,v,w), \qquad z = H(u,v,w). \tag{9}$$

Hence,

$$\begin{aligned} dx &= \frac{\partial F}{\partial u} du + \frac{\partial F}{\partial v} dv + \frac{\partial F}{\partial w} dw, \\ dy &= \frac{\partial G}{\partial u} du + \frac{\partial G}{\partial v} dv + \frac{\partial G}{\partial w} dw, \\ dz &= \frac{\partial H}{\partial u} du + \frac{\partial H}{\partial v} dv + \frac{\partial H}{\partial w} dw. \end{aligned} \tag{10}$$

Substituting in (4), we obtain

$$\begin{aligned} ds^2 = g_{11}\, du^2 + g_{22}\, dv^2 + g_{33}\, dw^2 + 2g_{12}\, du\, dv \\ + 2g_{13}\, du\, dw + 2g_{23}\, dv\, dw, \end{aligned} \tag{11}$$

where

$$\begin{aligned} g_{11} &= \left(\frac{\partial F}{\partial u}\right)^2 + \left(\frac{\partial G}{\partial u}\right)^2 + \left(\frac{\partial H}{\partial u}\right)^2; \\ g_{12} &= \frac{\partial F}{\partial u}\frac{\partial F}{\partial v} + \frac{\partial G}{\partial u}\frac{\partial G}{\partial v} + \frac{\partial H}{\partial u}\frac{\partial H}{\partial v}; \cdots. \end{aligned} \tag{12}$$

It can be shown that if the surfaces (8) are orthogonal, then $g_{mn} = 0$ if $m \neq n$. The proof is lengthy and we shall omit it since we are interested primarily in the actual expression for ds. In each special case we shall have direct evidence that ds is of the form (7).

For instance, in cylindrical coordinates $ds_\rho = d\rho$, $ds_\varphi = \rho d\varphi$, $ds_z = dz$; in spherical coordinates, $ds_r = dr$, $ds_\theta = r d\theta$, $ds_\varphi = r \sin\theta\, d\varphi$. The distance ds_θ is taken along a meridian of radius r and it is the product of r and the angle supported by the arc; ds_φ is taken along a circle of radius $r \sin\theta$ and the preceding formula follows. The coordinate lines are orthogonal and we obtain an expression of the form (7).

The physical character of space is determined by measurements of distances and hence by the corresponding *metrical form* (11). If, by some transformation of coordinates, the coefficients of the metrical form can be made constant, the space is said to be " flat "; otherwise, it is " curved." The words " flat " and " curved " are borrowed from the two-dimensional geometry of planes and curved surfaces and should not be taken too

literally, at least in so far as physical space is concerned. On the surface of a sphere of radius a, $ds^2 = a^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$; by transformation of coordinates this expression can be thrown into a great variety of forms but never into $ds^2 = du^2 + dv^2$. By careful measurements made *without getting outside the surface* it would be possible to establish that the geometry of the surface is different from that of a plane and that it is identical with the geometry of a sphere. Similarly, only measurements can disclose whether the physical space is flat — *pythagorean* might be a better term — or curved (nonpythagorean). Expressing θ and φ in terms of *three* variables, it is possible to choose the new variables in such a way that ds^2 will equal the sum of the squares of the differential elements; thus there is an abstract basis for asserting that the surface of a sphere may be immersed in a three-dimensional pythagorean space. Similarly a three-dimensional nonpythagorean space may be regarded as "immersed" in a pythagorean space of a larger number of dimensions.

The term flat or pythagorean two-dimensional space is not completely synonymous with the term "plane." The differential elements on cylindrical and other "developable" surfaces may be expressed in the simple form (4).

In any system of orthogonal coordinates the elements of area and volume are

$$\begin{aligned} dS_{uv} &= ds_u\, ds_v = e_1e_2\, du\, dv,\\ dS_{vw} &= ds_v\, ds_w = e_2e_3\, dv\, dw,\\ dS_{wu} &= ds_w\, ds_u = e_3e_1\, dw\, du;\\ d\tau &= ds_u\, ds_v\, ds_w = e_1e_2e_3\, du\, dv\, dw. \end{aligned} \tag{13}$$

It is these expressions that should be used in changing coordinates in surface and volume integrals.

3. *Calculation of gradient, divergence, curl, laplacian*

The definitions of the gradient of a scalar point function, and of the divergence and curl of a vector point function, as given in the preceding chapter, are independent of coordinate systems and are well adapted for the mathematical description of relations between physical quantities; but, in general, it would be awkward to have to depend on them for the actual calculation of these functions. It is more convenient to express their values in terms of the derivatives of the functions in question in various coordinate systems.

Grad V is a vector which can be resolved in any three mutually perpendicular directions; these can be taken as the u, v, w directions. The

components are the directional derivatives; hence,

$$\operatorname{grad}_u V = \frac{dV}{ds_u} = \frac{1}{e_1}\frac{\partial V}{\partial u}, \qquad \operatorname{grad}_v V = \frac{dV}{ds_v} = \frac{1}{e_2}\frac{\partial V}{\partial v},$$
$$\operatorname{grad}_w V = \frac{dV}{ds_w} = \frac{1}{e_3}\frac{\partial V}{\partial w}. \tag{14}$$

In order to obtain div F at some point P we surround P by an elementary coordinate cell like the one shown in Figure 8.2. The definition instructs us to divide the flux through the boundaries of the cell by the volume. For an elementary cell it is convenient to obtain the flux across the opposite faces as follows: consider the vw surface passing through P and suppose that the area intercepted by the cell is dS_{vw}; the flux across this area is $F_u\, dS_{vw}$; the rate of change of this flux as the vw surface is moved across the cell in the positive u direction is $D_u(F_u\, dS_{vw})$ and the differential change in the flux as dS_{vw} is moved from one face of the cell to the other in the positive u direction is $D_u(F_u\, dS_{vw})\, du$. Similarly we obtain the contributions of the two remaining pairs of faces; then

$$\begin{aligned}\operatorname{div} F &= \frac{D_u(F_u\, dS_{vw})\, du + D_v(F_v\, dS_{uw})\, dv + D_w(F_w\, dS_{uv})\, dw}{d\tau} \\ &= \frac{1}{e_1e_2e_3}\left[\frac{\partial}{\partial u}(e_2e_3F_u) + \frac{\partial}{\partial v}(e_3e_1F_v) + \frac{\partial}{\partial w}(e_1e_2F_w)\right].\end{aligned} \tag{15}$$

When deriving these equations it should be noted that, in taking the directional derivative D_u of $F_u\, dS_{vw} = F_u e_2 e_3\, dv\, dw$, the product $dv\, dw$ is constant; this is true because we have chosen coordinate surfaces as the boundaries of the cell.

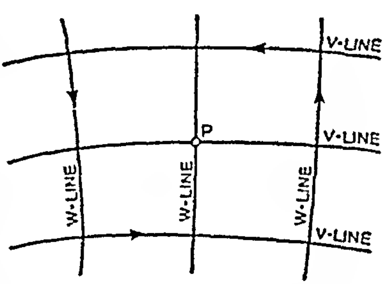

FIG. 8.3. An elementary curvilinear rectangle formed by the intersection of two pairs of neighboring coordinate lines lying in the same coordinate surface. Another pair of coordinate lines is shown passing through a point P.

To calculate $\operatorname{curl}_u F$ we consider the vw surface through P and surround the point by a differential rectangle of v and w lines, Figure 8.3. The circulation round this rectangle is to be divided by the area $dS_{vw} = e_2e_3\, dv\, dw$. The simplest way of computing the circulation is to consider the contributions from the pairs of opposite sides of the rectangle. Take the v-line passing through P; the line integral of F over that portion of this line which is intercepted by the rectangle is $F_v\, ds_v$; the rate of change of this in the positive w direction is $D_w(F_v\, ds_v)$; the change in the line

integral is $D_w(F_v\,ds_v)\,dw$; and this is the contribution to the *clockwise* circulation. Similarly $D_v(F_w\,ds_w)\,dv$ is the contribution to the *counterclockwise* circulation from the remaining two sides. Since $\text{curl}_u F$ is the *clockwise* circulation per unit area when the observer is looking in the positive u direction, that is, the counterclockwise circulation for the reader looking at Figure 8.3, we have

$$\begin{aligned}
\text{curl}_u F &= \frac{D_v(F_w\,ds_w)\,dv - D_w(F_v\,ds_v)\,dw}{dS_{vw}} \\
&= \frac{D_v(e_3F_w)\,dv\,dw - D_w(e_2F_v)\,dv\,dw}{e_2e_3\,dv\,dw} \qquad (16)\\
&= \frac{1}{e_2e_3}\left[\frac{\partial(e_3\,F_w)}{\partial v} - \frac{\partial(e_2F_v)}{\partial w}\right].
\end{aligned}$$

Permuting u, v, w and 1, 2, 3 cyclically we have

$$\begin{aligned}
\text{curl}_v F &= \frac{1}{e_3e_1}\left[\frac{\partial(e_1F_u)}{\partial w} - \frac{\partial(e_3F_w)}{\partial u}\right], \\
\text{curl}_w F &= \frac{1}{e_1e_2}\left[\frac{\partial(e_2F_v)}{\partial u} - \frac{\partial(e_1F_u)}{\partial v}\right].
\end{aligned} \qquad (17)$$

The laplacian $\Delta V = \text{div grad } V$; hence

$$\Delta V = \frac{1}{e_1e_2e_3}\left[\frac{\partial}{\partial u}\left(\frac{e_2e_3}{e_1}\frac{\partial V}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{e_3e_1}{e_2}\frac{\partial V}{\partial v}\right) + \frac{\partial}{\partial w}\left(\frac{e_1e_2}{e_3}\frac{\partial V}{\partial w}\right)\right]. \qquad (18)$$

4. *Transformation of coordinates*

Transformation of scalar functions from one system of coordinates to another is a problem of substitution. To express $V(x,y,z)$ in the u, v, w system, we substitute for x, y, z their expressions (9) in terms of the new coordinates. In transforming a vector function $F(x,y,z)$ we have to bear in mind that simple substitution would give us the components in the new coordinates but in the old directions — and we are naturally interested in the components fitting the new frame of reference.

Now F_u is the projection of F on the u direction and is equal to the sum of the projections of F_x, F_y, F_z; similarly in the case of F_v, F_w and therefore

$$\begin{aligned}
F_u &= F_x \cos(x,u) + F_y \cos(y,u) + F_z \cos(z,u), \\
F_v &= F_x \cos(x,v) + F_y \cos(y,v) + F_z \cos(z,v), \\
F_w &= F_x \cos(x,w) + F_y \cos(y,w) + F_z \cos(z,w).
\end{aligned} \qquad (19)$$

One of the coordinate systems happens to be cartesian by accident only;

the form of the equations of transformation is perfectly general and the coefficients are the cosines of the angles between the two sets of directions.

Frequently, these direction cosines can be obtained directly from geometric considerations. General expressions can be derived by transforming some simple vector and comparing the coefficients; the vector $\vec{ds}$ joining two infinitely close points is ideally suited for this purpose since its components ds_u, ds_v, ds_w are easily obtained when the equations relating the coordinate systems are known. Since $ds_u = e_1\,du$ and du is the total differential of $u = u(x,y,z)$, we have

$$\begin{aligned} ds_u &= e_1\,du = e_1\frac{\partial u}{\partial x}dx + e_1\frac{\partial u}{\partial y}dy + e_1\frac{\partial u}{\partial z}dz, \\ ds_v &= e_2\,dv = e_2\frac{\partial v}{\partial x}dx + e_2\frac{\partial v}{\partial y}dy + e_2\frac{\partial v}{\partial z}dz, \\ ds_w &= e_3\,dw = e_3\frac{\partial w}{\partial x}dx + e_3\frac{\partial w}{\partial y}dy + e_3\frac{\partial w}{\partial z}dz. \end{aligned} \tag{20}$$

Since $dx = as_x$, $dy = ds_y$, $dz = ds_z$ are the components of the same vector in the cartesian system, the coefficients are the direction cosines.

Sometimes it may be more convenient to express the cartesian components of $\vec{ds}$ in terms of the components of the other system:

$$\begin{aligned} dx &= \frac{\partial x}{\partial u}du + \frac{\partial x}{\partial v}dv + \frac{\partial x}{\partial w}dw = \frac{1}{e_1}\frac{\partial x}{\partial u}ds_u + \frac{1}{e_2}\frac{\partial x}{\partial v}ds_v + \frac{1}{e_3}\frac{\partial x}{\partial w}ds_w, \\ dy &= \frac{1}{e_1}\frac{\partial y}{\partial u}ds_u + \frac{1}{e_2}\frac{\partial y}{\partial v}ds_v + \frac{1}{e_3}\frac{\partial y}{\partial w}ds_w, \\ dz &= \frac{1}{e_1}\frac{\partial z}{\partial u}ds_u + \frac{1}{e_2}\frac{\partial z}{\partial v}ds_v + \frac{1}{e_3}\frac{\partial z}{\partial w}ds_w. \end{aligned} \tag{21}$$

From this set of equations we have

$$\cos(u,x) = \frac{1}{e_1}\frac{\partial x}{\partial u}, \qquad \cos(v,x) = \frac{1}{e_2}\frac{\partial x}{\partial v}, \cdots \tag{22}$$

while from (20) the corresponding expressions are

$$\cos(u,x) = e_1\frac{\partial u}{\partial x}, \qquad \cos(v,x) = e_2\frac{\partial v}{\partial x}, \cdots. \tag{23}$$

That there may be a real difference in the ease with which the results are obtained may be seen from (1). It is far easier to differentiate x, y, z with respect to r, θ, φ than vice versa. Incidentally equations (22) and (23) show that $\partial u/\partial x$ is not equal to the reciprocal of $\partial x/\partial u$.

Problems

1. Show that the cosines of the angles made by coordinate lines of any two orthogonal systems (u,v,w) and $(\hat{u},\hat{v},\hat{w})$ are

$$\cos(\hat{u},u) = (\hat{e}_1/e_1)\frac{\partial \hat{u}}{\partial u} = (e_1/\hat{e}_1)\frac{\partial u}{\partial \hat{u}},$$

$$\cos(\hat{u},v) = (\hat{e}_1/e_2)\frac{\partial \hat{u}}{\partial v} = (e_2/\hat{e}_1)\frac{\partial v}{\partial \hat{u}},\cdots.$$

2. Calculate the transformation cosines for the cartesian and spherical coordinate systems.

5. *Special coordinate systems*

Some of the most frequently used coordinate systems will now be described.

Cartesian coordinates. There are two systems, the right-handed as in Figure 8.1 and the left-handed. The left-handed system is obtained from the right-handed by reversing the positive direction of one or all of the coordinate axes. In applied mathematics the right-handed system enjoys general acceptance.

The expressions for the gradient, divergence, laplacian and curl are

$$\begin{aligned}
&\operatorname{grad}_x V = \frac{\partial V}{\partial x}, \qquad \operatorname{grad}_y V = \frac{\partial V}{\partial y}, \qquad \operatorname{grad}_z V = \frac{\partial V}{\partial z};\\
&\operatorname{div} F = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}; \qquad \Delta V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2};\\
&\operatorname{curl}_x F = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}, \qquad \operatorname{curl}_y F = \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x},\\
&\operatorname{curl}_z F = \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}.
\end{aligned} \tag{24}$$

Cylindrical coordinates. Figure 8.1 shows the most usual orientation of the cylindrical system with reference to the cartesian. The expressions for cylindrical coordinates (ρ,φ,z) in terms of cartesian are given in Section 1. The direction cosines are

	x	y	z
ρ	$\cos\varphi$	$\sin\varphi$	0
φ	$-\sin\varphi$	$\cos\varphi$	0
z	0	0	1

(25)

The fundamental metrical form and the expressions for the gradient, divergence, laplacian and curl are

$$\begin{aligned}
&ds^2 = d\rho^2 + \rho^2\, d\varphi^2 + dz^2;\\
&\mathrm{grad}_\rho V = \frac{\partial V}{\partial \rho}, \qquad \mathrm{grad}_\varphi V = \frac{\partial V}{\rho\, \partial \varphi}, \qquad \mathrm{grad}_z V = \frac{\partial V}{\partial z};\\
&\mathrm{div}\, F = \frac{1}{\rho}\frac{\partial(\rho F_\rho)}{\partial \rho} + \frac{\partial F_\varphi}{\rho\, \partial \varphi} + \frac{\partial F_z}{\partial z};\\
&\Delta V = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial V}{\partial \rho}\right) + \frac{\partial^2 V}{\rho^2\, \partial \varphi^2} + \frac{\partial^2 V}{\partial z^2};\\
&\mathrm{curl}_\rho F = \frac{\partial F_z}{\rho\, \partial \varphi} - \frac{\partial F_\varphi}{\partial z}, \qquad \mathrm{curl}_\varphi F = \frac{\partial F_\rho}{\partial z} - \frac{\partial F_z}{\partial \rho},\\
&\mathrm{curl}_z F = \frac{1}{\rho}\left[\frac{\partial(\rho F_\varphi)}{\partial \rho} - \frac{\partial F_\rho}{\partial \varphi}\right].
\end{aligned} \tag{26}$$

Spherical coordinates. Figure 8.1 shows spherical coordinates in their relation to cartesian and cylindrical coordinates. Expressions for r, θ, φ are given in Section 1. The fundamental metrical form and other formulas are:

$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2;$$

	x	y	z
r	$\sin\theta\cos\varphi$	$\sin\theta\sin\varphi$	$\cos\theta$
θ	$\cos\theta\cos\varphi$	$\cos\theta\sin\varphi$	$-\sin\theta$
φ	$-\sin\varphi$	$\cos\varphi$	0

$$\begin{aligned}
&\mathrm{grad}_r V = \frac{\partial V}{\partial r}, \qquad \mathrm{grad}_\theta V = \frac{\partial V}{r\, \partial \theta}, \qquad \mathrm{grad}_\varphi V = \frac{\partial V}{r\sin\theta\, \partial \varphi};\\
&\mathrm{div}\, F = \frac{1}{r^2}\frac{\partial(r^2 F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\theta)}{\partial \theta} + \frac{\partial F_\varphi}{r\sin\theta\, \partial \varphi};\\
&\Delta V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial \varphi^2};\\
&\mathrm{curl}_r F = \frac{1}{r\sin\theta}\left[\frac{\partial(\sin\theta F_\varphi)}{\partial \theta} - \frac{\partial F_\theta}{\partial \varphi}\right],\\
&\mathrm{curl}_\theta F = \frac{1}{r\sin\theta}\left[\frac{\partial F_r}{\partial \varphi} - \sin\theta\frac{\partial(rF_\varphi)}{\partial r}\right],\\
&\mathrm{curl}_\varphi F = \frac{1}{r}\left[\frac{\partial(rF_\theta)}{\partial r} - \frac{\partial F_r}{\partial \theta}\right].
\end{aligned} \tag{27}$$

Elliptic cylinder coordinates. There are two related forms of elliptic cylinder coordinates (ξ,ϑ,z) and (u,v,z). Their connection with cartesian coordinates may be expressed as a transformation of complex variables*

$$\begin{aligned} x + iy &= f\cosh(\xi + i\vartheta); \\ x &= f\cosh\xi\cos\vartheta = fuv, \qquad u = \cosh\xi, \qquad v = \cos\vartheta; \\ y &= f\sinh\xi\sin\vartheta = f\sqrt{(u^2-1)(1-v^2)}; \\ ds^2 &= \tfrac{1}{2}f^2(\cosh 2\xi - \cos 2\vartheta)(d\xi^2 + d\vartheta^2) + dz^2 \\ &= f^2(u^2 - v^2)\left(\frac{du^2}{u^2-1} + \frac{dv^2}{1-v^2}\right) + dz^2; \\ \rho^2 &= \tfrac{1}{2}f^2(\cosh 2\xi + \cos 2\vartheta) = f^2(u^2 + v^2 - 1); \\ \tan\varphi &= \tanh\xi\tan\vartheta. \end{aligned} \tag{28}$$

The coordinate surfaces are confocal elliptic and hyperbolic cylinders

$$\frac{x^2}{f^2\cosh^2\xi} + \frac{y^2}{f^2\sinh^2\xi} = 1, \qquad \frac{x^2}{f^2\cos^2\vartheta} - \frac{y^2}{f^2\sin^2\vartheta} = 1, \tag{29}$$

and planes z = constant. The geometric meaning of elliptic coordinates ξ, ϑ and u, v is illustrated in Figure 8.4a and may be stated as follows:

$OF = f$ is the distance from the center to a common focus F,
$OA = f\cosh\xi = fu$ is the semi-major axis of an ellipse ξ = constant,
$OB = f\sinh\xi = f\sqrt{u^2-1}$ is the semi-minor axis of the ellipse,
$e = 1/u = \operatorname{sech}\xi$ is the eccentricity of the ellipse,
$OC = f\cos\vartheta = fv$ is the real semi-axis of a hyperbola ϑ = constant,
$CD = f\sin\vartheta = f\sqrt{1-v^2}$ is the imaginary semi-axis of the hyperbola,
$e' = 1/v = \sec\vartheta$ is the eccentricity of the hyperbola,
$\angle\, COD = \vartheta$ is the angle between the x-axis and an asymptote to the hyperbola; it is also the eccentric angle for the point P on the ellipse. (In order to construct the eccentric angle for any given point on a given ellipse draw a circle of radius $OA = f\cosh\xi$, equal to the semi-major axis; extend the ordinate through the given point P to the intersection E with the circle; the angle AOE is the eccentric angle ϑ since $x = (f\cosh\xi)\cos\vartheta$ is the abscissa of P.)

Thus $u = \cosh\xi$ and $v = \cos\vartheta$ are the reciprocals of the eccentricities of

* To obtain ds^2, take the differential of $x + iy$ and multiply it by its conjugate.

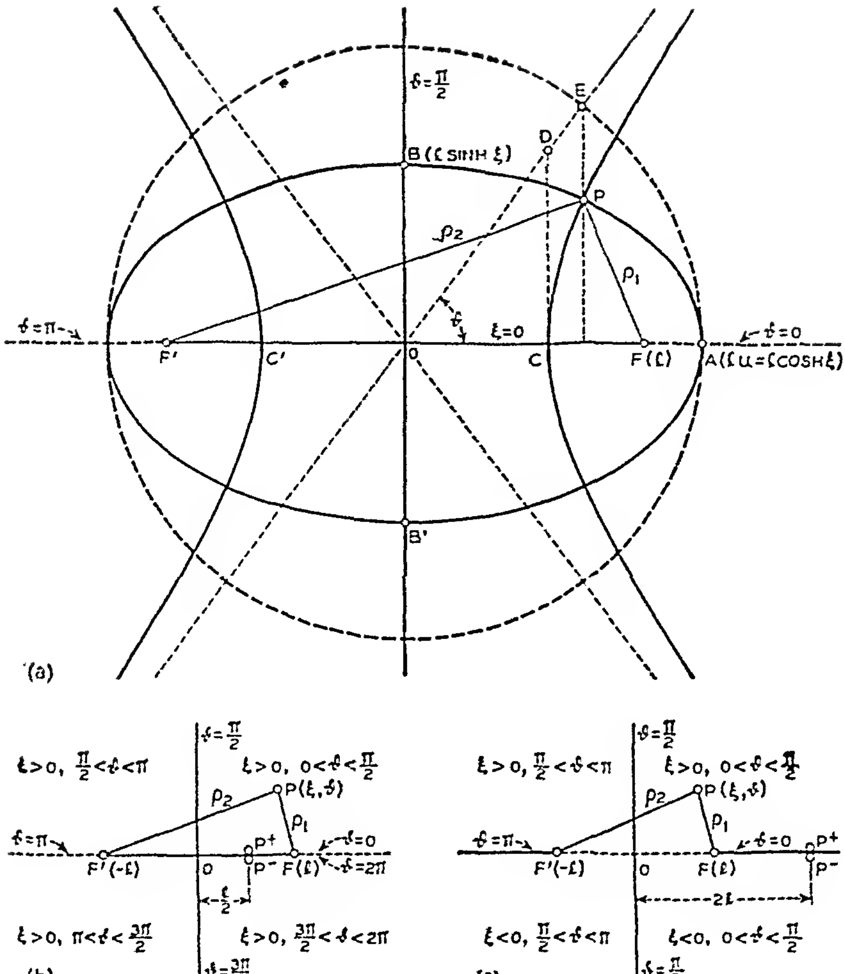

FIG. 8.4a. Elliptic coordinates (u,v) or (ξ,ϑ). In the xy-plane, the lines of constant u and ξ are confocal ellipses; the lines of constant v and ϑ are confocal hyperbolas. The distance $2l$ between the foci is a characteristic constant of a particular system of elliptic coordinates. The values of the elliptic coordinates may be obtained from $u = (\rho_2 + \rho_1)/2l$ and $v = (\rho_2 - \rho_1)/2l$; $\xi = \cosh^{-1}[(\rho_2 + \rho_1)/2l]$ and $\vartheta = \cos^{-1}[(\rho_2 - \rho_1)/2l]$.

FIG. 8.4b. In the elliptic coordinate system of the *first kind* ξ is positive for all points and ϑ lies in the interval $(0,2\pi)$. The points P^+ and P^-, on the upper and lower sides of the segment $F'F$ joining the focal points, have different ϑ coordinates, $\cos^{-1}(x/l)$ and $2\pi - \cos^{-1}(x/l)$ respectively. Thus for the point, whose cartesian coordinates are $x = l/2$, $y = 0$, we have $\vartheta = \pi/3$ on the upper side and $\vartheta = 5\pi/3$ on the lower side.

FIG. 8.4c. In the elliptic coordinate system of the *second kind* ξ may be either positive or negative, while ϑ is restricted to the interval $(0,\pi)$. The x-axis, exclusive of the segment $F'F$ joining the focal points, has two sides. The ξ coordinates of points, P^+, on the upper side are positive; on the lower side negative. Thus if the cartesian coordinates of a point P are $x = 2l$, $y = 0$, then for the upper side $\xi = 1.317$ and for the lower side $\xi = -1.317$.

the confocal ellipse and hyperbola passing through the point (x,y). Also

$$u = \cosh \xi = \frac{\rho_2 + \rho_1}{2l}, \qquad v = \cos \vartheta = \frac{\rho_2 - \rho_1}{2l}, \tag{30}$$

where ρ_1 and ρ_2 are the distances between $P(u,v)$ and the foci.

If the cartesian coordinates are given, the elliptic coordinates may be found from

$$u^2 = \cosh^2 \xi = \frac{1}{2}\left(1 + \frac{x^2 + y^2}{l^2}\right) \pm \frac{1}{2}\sqrt{\left(1 + \frac{x^2 + y^2}{l^2}\right)^2 - \frac{4x^2}{l^2}}, \tag{31}$$

$$\cos \vartheta = \frac{x}{l \cosh \xi}, \qquad \sin \vartheta = \frac{y}{l \sinh \xi}. \tag{32}$$

If y is different from zero, the upper sign must be used to make u consistent with its geometric meaning; u is the ratio of the major axis and focal distance of the ellipse and must not be smaller than unity. The negative sign leads to a value smaller than unity, unless $y = 0$ in which case $u = 1$ and $\xi = 0$ is the second solution. The solution corresponds to points on the segment $F'F$; thus $\xi = 0$ is the equation of the segment $F'F$ in elliptic coordinates. Various points on $F'F$ are specified by $\cos \vartheta = x/l$.

Two alternatives now present themselves since neither ξ nor ϑ is uniquely determined by x and y. Thus $\cosh(-\xi) = \cosh \xi$ and $\cos(2\pi - \vartheta) = \cos \vartheta$; hence, when ξ is one solution of $\cosh \xi =$ a given value, then $-\xi$ is another possible solution. And if ϑ is one solution of $\cos \vartheta =$ a given value, then $2\pi - \vartheta$ is another solution. Suppose that we agree to take only positive values of ξ. The denominators in (32) will then be positive and in order to specify all points (x,y) by different pairs (ξ,ϑ) of elliptic coordinates we must let ϑ vary from 0 to 2π (or $-\pi$ to π). In this case ϑ is between 0 and π for points above the x-axis and between π and 2π for points below the x-axis. Therefore, one and the same point on the segment $F'F$ joining the foci of the coordinate system will be given by two different ϑ-coordinates, Figure 8.4b. We can take advantage of this result whenever it is convenient to distinguish between two sides of the segment $F'F$, as for instance in problems involving liquid flow in the presence of an infinite rigid strip or in electrical problems involving a metal strip, where it is obviously necessary to distinguish between two sides of the strip even if, for mathematical simplicity, we assume that the thickness of the strip is zero.

In the elliptic coordinate system of the *second kind*, Figure 8.4c, ξ is permitted to have negative values as well as positive and ϑ is restricted to the interval $(0,\pi)$. This time the x-axis, *exclusive* of the segment $F'F$, has

two sides; and the coordinate system is well adapted to problems involving plane screens with a slot.

The expressions for the gradient, divergence, laplacian and curl are:

$$\text{grad}_\xi V = \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial V}{\partial \xi},$$

$$\text{grad}_\vartheta V = \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial V}{\partial \vartheta}, \qquad \text{grad}_z V = \frac{\partial V}{\partial z};$$

$$\text{div}\, F = \frac{\sqrt{2}}{f(\cosh 2\xi - \cos 2\vartheta)}\left[\frac{\partial}{\partial \xi}(F_\xi\sqrt{\cosh 2\xi - \cos 2\vartheta}) + \frac{\partial}{\partial \vartheta}(F_\vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta})\right] + \frac{\partial F_z}{\partial z};$$

$$\Delta V = \frac{2}{f^2(\cosh 2\xi - \cos 2\vartheta)}\left(\frac{\partial^2 V}{\partial \xi^2} + \frac{\partial^2 V}{\partial \vartheta^2}\right) + \frac{\partial^2 V}{\partial z^2}; \tag{33}$$

$$\text{curl}_\xi F = \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial F_z}{\partial \vartheta} - \frac{\partial F_\vartheta}{\partial z},$$

$$\text{curl}_\vartheta F = \frac{\partial F_\xi}{\partial z} - \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial F_z}{\partial \xi},$$

$$\text{curl}_z F = \frac{\sqrt{2}}{f(\cosh 2\xi - \cos 2\vartheta)}\left[\frac{\partial}{\partial \xi}(F_\vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta}) - \frac{\partial}{\partial \vartheta}(F_\xi\sqrt{\cosh 2\xi - \cos 2\vartheta})\right].$$

Prolate spheroidal coordinates. Prolate spheroidal coordinates are defined by

$$z + i\rho = f\cosh(\xi + i\vartheta), \qquad 0 \le \xi < \infty, \qquad 0 \le \vartheta \le \pi;$$

$$z = f\cosh\xi\cos\vartheta, \qquad \rho = f\sinh\xi\sin\vartheta; \tag{34}$$

$$ds^2 = \tfrac{1}{2}f^2[(\cosh 2\xi - \cos 2\vartheta)(d\xi^2 + d\vartheta^2)] + f^2\sinh^2\xi\sin^2\vartheta\, d\varphi^2.$$

The z and ρ coordinates correspond respectively to x and y in Figure 8.4a. The coordinate surfaces are prolate spheroids, hyperboloids of revolution (about the horizontal axis in the figure) and axial planes. The segment $F'F$ is the limiting spheroid; its equation is $\xi = 0$. The positive z space is given by $0 < \vartheta < \pi/2$, the negative z space by $\pi/2 < \vartheta < \pi$.

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_\xi V = \frac{\sqrt{2}}{\ell\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial V}{\partial \xi},$$

$$\operatorname{grad}_\vartheta V = \frac{\sqrt{2}}{\ell\sqrt{\cosh 2\xi - \cos 2\vartheta}} \frac{\partial V}{\partial \vartheta},$$

$$\operatorname{grad}_\varphi V = \frac{1}{\ell \sinh \xi \sin \vartheta} \frac{\partial V}{\partial \varphi};$$

$$\operatorname{div} F = \frac{\sqrt{2}}{\ell(\cosh 2\xi - \cos 2\vartheta)}\left[\frac{1}{\sinh \xi}\frac{\partial}{\partial \xi}\left(\sinh \xi\sqrt{\cosh 2\xi - \cos 2\vartheta}\, F_\xi\right)\right.$$
$$\left. + \frac{1}{\sin \vartheta}\frac{\partial}{\partial \vartheta}\left(\sin \vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta}\, F_\vartheta\right)\right] + \frac{1}{\ell \sinh \xi \sin \vartheta}\frac{\partial F_\varphi}{\partial \varphi};$$

$$\Delta V = \frac{2}{\ell^2(\cosh 2\xi - \cos 2\vartheta)}\left[\frac{1}{\sinh \xi}\frac{\partial}{\partial \xi}\left(\sinh \xi \frac{\partial V}{\partial \xi}\right)\right. \tag{35}$$
$$\left. + \frac{1}{\sin \vartheta}\frac{\partial}{\partial \vartheta}\left(\sin \vartheta \frac{\partial V}{\partial \vartheta}\right)\right] + \frac{1}{\ell^2 \sinh^2 \xi \sin^2 \vartheta}\frac{\partial V^2}{\partial \varphi^2};$$

$$\operatorname{curl}_\xi F = \frac{\sqrt{2}}{\ell \sin \vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial}{\partial \vartheta}(\sin \vartheta F_\varphi) - \frac{1}{\ell \sinh \xi \sin \vartheta}\frac{\partial F_\vartheta}{\partial \varphi},$$

$$\operatorname{curl}_\vartheta F = \frac{1}{\ell \sinh \xi \sin \vartheta}\frac{\partial F_\xi}{\partial \varphi} - \frac{\sqrt{2}}{\ell \sinh \xi\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial}{\partial \xi}(\sinh \xi F_\varphi),$$

$$\operatorname{curl}_\varphi F = \frac{\sqrt{2}}{\ell(\cosh 2\xi - \cos 2\vartheta)}\left[\frac{\partial}{\partial \xi}\left(\sqrt{\cosh 2\xi - \cos 2\vartheta}\, F_\vartheta\right)\right.$$
$$\left. - \frac{\partial}{\partial \vartheta}\left(\sqrt{\cosh 2\xi - \cos 2\vartheta}\, F_\xi\right)\right].$$

Oblate spheroidal coordinates. Oblate spheroidal coordinates are defined by

$$\rho + iz = \ell \cosh(\xi + i\vartheta), \qquad 0 \le \xi < \infty, \qquad -\pi/2 \le \vartheta \le \pi/2;$$
$$\rho = \ell \cosh \xi \cos \vartheta, \qquad z = \ell \sinh \xi \sin \vartheta; \tag{36}$$
$$ds^2 = \tfrac{1}{2}\ell^2[(\cosh 2\xi - \cos 2\vartheta)(d\xi^2 + d\vartheta^2)] + \ell^2 \cosh^2 \xi \cos^2 \vartheta \, d\varphi^2.$$

The ρ and z coordinates correspond respectively to x and y in Figure 8.4a. The coordinate surfaces are oblate spheroids, hyperboloids of revolution (about the vertical axis in the figure) and axial planes. Positive and

negative z spaces correspond to positive and negative ϑ. The disc of radius OF is the limiting oblate spheroid; its equation is $\xi = 0$; the upper side is given by positive ϑ and the lower by negative ϑ.

In problems involving a circular hole in an infinite plane ϑ is restricted to the interval $(0,\pi/2)$ and ξ is permitted to have negative values. The equation of the perforated plane is $\vartheta = 0$; and each point in the hole is given by a unique pair of coordinates.

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_\xi V = \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial V}{\partial \xi},$$

$$\operatorname{grad}_\vartheta V = \frac{\sqrt{2}}{f\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial V}{\partial \vartheta},$$

$$\operatorname{grad}_\varphi V = \frac{1}{f\cosh\xi\cos\vartheta}\frac{\partial V}{\partial \varphi};$$

$$\begin{aligned}\operatorname{div} F = \frac{\sqrt{2}}{f(\cosh 2\xi - \cos 2\vartheta)}&\left[\frac{1}{\cosh\xi}\frac{\partial}{\partial\xi}\left(\cosh\xi\sqrt{\cosh 2\xi - \cos 2\vartheta}\,F_\xi\right)\right.\\ &\left.+\frac{1}{\cos\vartheta}\frac{\partial}{\partial\vartheta}\left(\cos\vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta}\,F_\vartheta\right)\right]\\ &+\frac{1}{f\cosh\xi\cos\vartheta}\frac{\partial F_\varphi}{\partial\varphi};\end{aligned}$$

$$\begin{aligned}\Delta V = \frac{2}{f^2(\cosh 2\xi - \cos 2\vartheta)}&\left[\frac{1}{\cosh\xi}\frac{\partial}{\partial\xi}\left(\cosh\xi\frac{\partial V}{\partial\xi}\right)\right.\\ &\left.+\frac{1}{\cos\vartheta}\frac{\partial}{\partial\vartheta}\left(\cos\vartheta\frac{\partial V}{\partial\vartheta}\right)\right]+\frac{1}{f^2\cosh^2\xi\cos^2\vartheta}\frac{\partial^2 V}{\partial\varphi^2};\end{aligned}$$

$$\operatorname{curl}_\xi F = \frac{\sqrt{2}}{f\cos\vartheta\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial}{\partial\vartheta}(\cos\vartheta F_\varphi) - \frac{1}{f\cosh\xi\cos\vartheta}\frac{\partial F_\vartheta}{\partial\varphi},$$

$$\operatorname{curl}_\vartheta F = \frac{1}{f\cosh\xi\cos\vartheta}\frac{\partial F_\xi}{\partial\varphi} - \frac{\sqrt{2}}{f\cosh\xi\sqrt{\cosh 2\xi - \cos 2\vartheta}}\frac{\partial}{\partial\xi}(\cosh\xi F_\varphi),$$

$$\begin{aligned}\operatorname{curl}_\varphi F = \frac{\sqrt{2}}{f(\cosh 2\xi - \cos 2\vartheta)}&\left[\frac{\partial}{\partial\xi}\left(\sqrt{\cosh 2\xi - \cos 2\vartheta}\,F_\vartheta\right)\right.\\ &\left.-\frac{\partial}{\partial\vartheta}\left(\sqrt{\cosh 2\xi - \cos 2\vartheta}\,F_\xi\right)\right].\end{aligned}$$

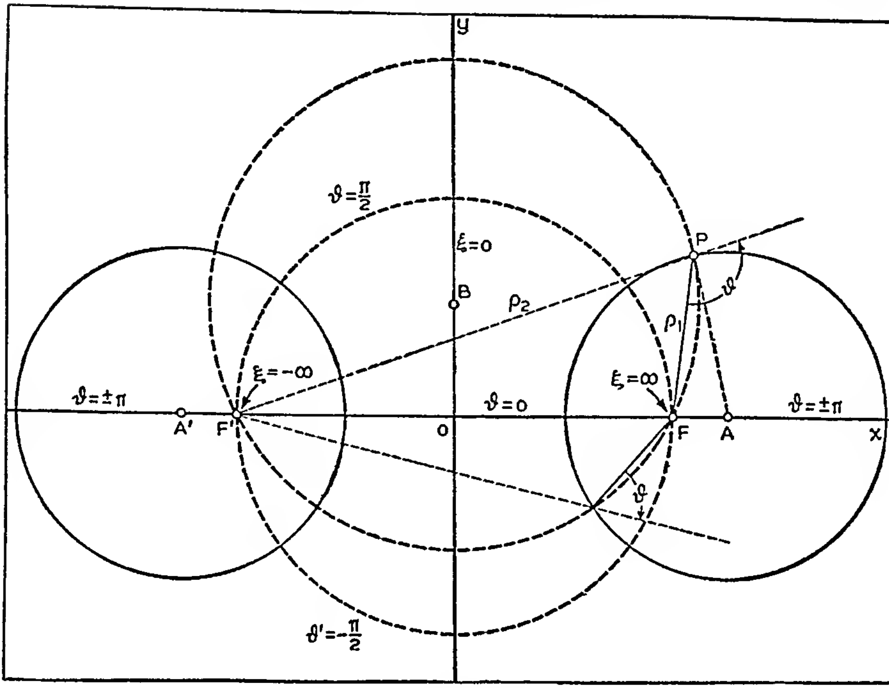

FIG. 8.5. Bipolar coordinate system in the xy-plane, or a z-section of the biaxial coordinate system in space.

Biaxial coordinates.

Biaxial coordinates are defined as follows:

$$x + iy = \ell \tanh \tfrac{1}{2}(\xi + i\vartheta), \qquad \xi + i\vartheta = \log \frac{\ell + x + iy}{\ell - x - iy};$$

$$x = \ell \frac{\sinh \xi}{\cosh \xi + \cos \vartheta}, \qquad y = \ell \frac{\sin \vartheta}{\cosh \xi + \cos \vartheta};$$

$$\xi = \frac{1}{2} \log \frac{(\ell + x)^2 + y^2}{(\ell - x)^2 + y^2}, \qquad \vartheta = \tan^{-1} \frac{y}{\ell + x} + \tan^{-1} \frac{y}{\ell - x};$$

$$\rho^2 = x^2 + y^2 = \ell^2 \frac{\cosh \xi - \cos \vartheta}{\cosh \xi + \cos \vartheta}; \qquad -\infty < \xi < \infty;$$

$$ds^2 = \frac{\ell^2 (d\xi^2 + d\vartheta^2)}{(\cosh \xi + \cos \vartheta)^2} + dz^2; \qquad -\pi \leq \vartheta \leq \pi.$$

Coordinate surfaces are circular cylinders passing through a pair of parallel *focal lines*, distance 2ℓ apart, an orthogonal family of circular cylinders, and planes perpendicular to the generators of these cylinders. Figure 8.5 shows a cross section of the cylinders by one of these planes. Focal lines

are limiting cylinders $\xi = \pm\infty$. The equation of the yz-plane is $\xi = 0$. The equation of the strip of the xz-plane between the focal lines is $\vartheta = 0$; the equation of the remainder of the xz-plane is $\vartheta = \pi$ (or $-\pi$). The equation of the coordinate cylinder coaxial with the z-axis is $\vartheta = \pm\pi/2$, the plus sign corresponding to one half and the minus sign to the other half.

The position of the center and the radius of a coordinate cylinder $\xi =$ constant are given by

$$x = \ell \coth \xi \qquad R = \ell \,|\operatorname{csch} \xi\,| .$$

The position of the center and the radius of a coordinate cylinder $\vartheta =$ constant are given by

$$y = -\ell \cot \vartheta, \qquad R = \ell \,|\csc \vartheta\,| .$$

The coordinate $\xi = \log(\rho_2/\rho_1)$, and ϑ is the angle through which we have to swing PF in Figure 8.5 to make it coincide with the continuation of $F'P$.

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_\xi V = \frac{1}{\ell}(\cosh \xi + \cos \vartheta)\frac{\partial V}{\partial \xi},$$

$$\operatorname{grad}_\vartheta V = \frac{1}{\ell}(\cosh \xi + \cos \vartheta)\frac{\partial V}{\partial \vartheta}, \qquad \operatorname{grad}_z V = \frac{\partial V}{\partial z};$$

$$\operatorname{div} F = \frac{1}{\ell}(\cosh \xi + \cos \vartheta)^2\left[\frac{\partial}{\partial \xi}\frac{F_\xi}{\cosh \xi + \cos \vartheta} + \frac{\partial}{\partial \vartheta}\frac{F_\vartheta}{\cosh \xi + \cos \vartheta}\right] + \frac{\partial F_z}{\partial z};$$

$$\Delta V = \frac{1}{\ell^2}(\cosh \xi + \cos \vartheta)^2\left(\frac{\partial^2 V}{\partial \xi^2} + \frac{\partial^2 V}{\partial \vartheta^2}\right) + \frac{\partial^2 V}{\partial z^2};$$

$$\operatorname{curl}_\xi F = \frac{1}{\ell}(\cosh \xi + \cos \vartheta)\frac{\partial F_z}{\partial \vartheta} - \frac{\partial F_\vartheta}{\partial z},$$

$$\operatorname{curl}_\vartheta F = \frac{\partial F_\xi}{\partial z} - \frac{1}{\ell}(\cosh \xi + \cos \vartheta)\frac{\partial F_z}{\partial \xi},$$

$$\operatorname{curl}_z F = \frac{1}{\ell}(\cosh \xi + \cos \vartheta)^2\left[\frac{\partial}{\partial \xi}\frac{F_\vartheta}{\cosh \xi + \cos \vartheta} - \frac{\partial}{\partial \vartheta}\frac{F_\xi}{\cosh \xi + \cos \vartheta}\right].$$

Toroidal coordinates. To obtain toroidal coordinates, we shall call the vertical axis in Figure 8.5 the z-axis and consider the figure of revolution about it. The circles with their centers on the z-axis will form spheres;

the circles with their centers on the ρ-axis (the former x-axis) will form toruses. The defining equations for toroidal coordinates are:

$$\rho + iz = \ell \tanh \tfrac{1}{2}(\xi + i\vartheta), \qquad \xi + i\vartheta = \log \frac{\ell + \rho + iz}{\ell - \rho - iz};$$

$$\rho = \ell \frac{\sinh \xi}{\cosh \xi + \cos \vartheta}, \qquad z = \ell \frac{\sin \vartheta}{\cosh \xi + \cos \vartheta};$$

$$\xi = \frac{1}{2} \log \frac{(\ell + \rho)^2 + z^2}{(\ell - \rho)^2 + z^2}, \qquad \vartheta = \tan^{-1} \frac{z}{\ell + \rho} + \tan^{-1} \frac{z}{\ell - \rho};$$

$$r^2 = \rho^2 + z^2 = \ell^2 \frac{\cosh \xi - \cos \vartheta}{\cosh \xi + \cos \vartheta}; \qquad 0 \le \xi < \infty;$$

$$ds^2 = \frac{\ell^2 (d\xi^2 + d\vartheta^2 + \sinh^2 \xi \, d\varphi^2)}{(\cosh \xi + \cos \vartheta)^2}; \qquad -\pi \le \vartheta \le \pi.$$

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_\xi V = \frac{1}{\ell} (\cosh \xi + \cos \vartheta) \frac{\partial V}{\partial \xi}, \qquad \operatorname{grad}_\vartheta V = \frac{1}{\ell} (\cosh \xi + \cos \vartheta) \frac{\partial V}{\partial \vartheta},$$

$$\operatorname{grad}_\varphi V = \frac{1}{\ell} \frac{\cosh \xi + \cos \vartheta}{\sinh \xi} \frac{\partial V}{\partial \varphi};$$

$$\operatorname{div} F = \frac{(\cosh \xi + \cos \vartheta)^3}{\ell \sinh \xi} \left[\frac{\partial}{\partial \xi} \frac{F_\xi \sinh \xi}{(\cosh \xi + \cos \vartheta)^2} + \frac{\partial}{\partial \vartheta} \frac{F_\vartheta \sinh \xi}{(\cosh \xi + \cos \vartheta)^2} \right] + \frac{\cosh \xi + \cos \vartheta}{\ell \sinh \xi} \frac{\partial F_\varphi}{\partial \varphi};$$

$$\Delta V = \frac{(\cosh \xi + \cos \vartheta)^3}{\ell^2 \sinh \xi} \left[\frac{\partial}{\partial \xi} \left(\frac{\sinh \xi}{\cosh \xi + \cos \vartheta} \frac{\partial V}{\partial \xi} \right) + \frac{\partial}{\partial \vartheta} \left(\frac{\sinh \xi}{\cosh \xi + \cos \vartheta} \frac{\partial V}{\partial \vartheta} \right) \right] + \frac{(\cosh \xi + \cos \vartheta)^2}{\ell^2 \sinh^2 \xi} \frac{\partial^2 V}{\partial \varphi^2};$$

$$\operatorname{curl}_\xi F = \frac{1}{\ell} (\cosh \xi + \cos \vartheta)^2 \frac{\partial}{\partial \vartheta} \frac{F_\varphi}{\cosh \xi + \cos \vartheta} - \frac{\cosh \xi + \cos \vartheta}{\ell \sinh \xi} \frac{\partial F_\vartheta}{\partial \varphi},$$

$$\operatorname{curl}_\vartheta F = \frac{\cosh \xi + \cos \vartheta}{\ell \sinh \xi} \frac{\partial F_\xi}{\partial \varphi} - \frac{(\cosh \xi + \cos \vartheta)^2}{\ell \sinh \xi} \frac{\partial}{\partial \xi} \frac{F_\varphi \sinh \xi}{\cosh \xi + \cos \vartheta},$$

$$\operatorname{curl}_\varphi F = \frac{1}{\ell} (\cosh \xi + \cos \vartheta)^2 \left[\frac{\partial}{\partial \xi} \frac{F_\vartheta}{\cosh \xi + \cos \vartheta} - \frac{\partial}{\partial \vartheta} \frac{F_\xi}{\cosh \xi + \cos \vartheta} \right].$$

Bipolar coordinates. Bipolar coordinates are obtained by taking the horizontal axis in Figure 8.5 as the z-axis and making it the axis of revolution. The circles with their centers on the horizontal axis will form spheres. The defining equations are:

$$z + i\rho = f \tanh \tfrac{1}{2}(\xi + i\vartheta), \qquad \xi + i\vartheta = \log \frac{f + z + i\rho}{f - z - i\rho};$$

$$z = f \frac{\sinh \xi}{\cosh \xi + \cos \vartheta}, \qquad \rho = f \frac{\sin \vartheta}{\cosh \xi + \cos \vartheta};$$

$$\xi = \frac{1}{2} \log \frac{(f + z)^2 + \rho^2}{(f - z)^2 + \rho^2}, \qquad \vartheta = \tan^{-1} \frac{\rho}{f + z} + \tan^{-1} \frac{\rho}{f - z}:$$

$$r^2 = \rho^2 + z^2 = f^2 \frac{\cosh \xi - \cos \vartheta}{\cosh \xi + \cos \vartheta}; \qquad -\infty < \xi < \infty;$$

$$ds^2 = \frac{f^2(d\xi^2 + d\vartheta^2 + \sin^2 \vartheta \, d\varphi^2)}{(\cosh \xi + \cos \vartheta)^2}; \qquad 0 \leq \vartheta \leq \pi.$$

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_\xi V = \frac{1}{f} (\cosh \xi + \cos \vartheta) \frac{\partial V}{\partial \xi}, \qquad \operatorname{grad}_\vartheta V = \frac{1}{f} (\cosh \xi + \cos \vartheta) \frac{\partial V}{\partial \vartheta},$$

$$\operatorname{grad}_\varphi V = \frac{1}{f} \frac{\cosh \xi + \cos \vartheta}{\sin \vartheta} \frac{\partial V}{\partial \varphi};$$

$$\operatorname{div} F = \frac{(\cosh \xi + \cos \vartheta)^3}{f \sin \vartheta} \left[\frac{\partial}{\partial \xi} \frac{F_\xi \sin \vartheta}{(\cosh \xi + \cos \vartheta)^2} + \frac{\partial}{\partial \vartheta} \frac{F_\vartheta \sin \vartheta}{(\cosh \xi + \cos \vartheta)^2} \right] + \frac{\cosh \xi + \cos \vartheta}{f \sin \vartheta} \frac{\partial F_\varphi}{\partial \varphi};$$

$$\Delta V = \frac{(\cosh \xi + \cos \vartheta)^3}{f^2 \sin^2 \vartheta} \left[\frac{\partial}{\partial \xi} \left(\frac{\sin \vartheta}{\cosh \xi + \cos \vartheta} \frac{\partial V}{\partial \xi} \right) + \frac{\partial}{\partial \vartheta} \left(\frac{\sin \vartheta}{\cosh \xi + \cos \vartheta} \frac{\partial V}{\partial \vartheta} \right) \right] + \frac{(\cosh \xi + \cos \vartheta)^2}{f^2 \sin^2 \vartheta} \frac{\partial^2 V}{\partial \varphi^2};$$

$$\operatorname{curl}_\xi F = \frac{(\cosh \xi + \cos \vartheta)^2}{f \sin \vartheta} \frac{\partial}{\partial \vartheta} \frac{F_\varphi \sin \vartheta}{\cosh \xi + \cos \vartheta} - \frac{\cosh \xi + \cos \vartheta}{f \sin \vartheta} \frac{\partial F_\vartheta}{\partial \varphi},$$

$$\operatorname{curl}_\vartheta F = \frac{\cosh \xi + \cos \vartheta}{f \sin \vartheta} \frac{\partial F_\xi}{\partial \varphi} - \frac{1}{f} (\cosh \xi + \cos \vartheta)^2 \frac{\partial}{\partial \xi} \frac{F_\varphi}{\cosh \xi + \cos \vartheta},$$

$$\operatorname{curl}_\varphi F = \frac{1}{f} (\cosh \xi + \cos \vartheta)^2 \left[\frac{\partial}{\partial \xi} \frac{F_\vartheta}{\cosh \xi + \cos \vartheta} - \frac{\partial}{\partial \vartheta} \frac{F_\xi}{\cosh \xi + \cos \vartheta} \right].$$

Parabolic cylinder coordinates. Parabolic coordinates in a plane are defined as follows

$$x + iy = \ell(u + iv)^2, \qquad x = \ell(u^2 - v^2), \qquad y = 2\ell uv.$$

The lines u = constant form a family of parabolas, concave to the left; these are shown as dotted lines in Figure 8.6. The lines v = constant

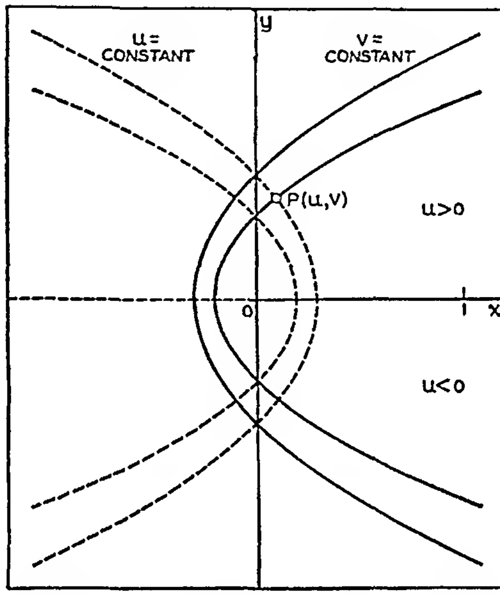

FIG. 8.6. Parabolic coordinates in the xy-plane, or a z-section of parabolic cylinder coordinates.

form an orthogonal family of parabolas, concave to the right; these are shown as solid lines. The corresponding cartesian equations are:

$$y^2 = 4\ell^2u^2\left(u^2 - \frac{x}{\ell}\right), \qquad y^2 = 4\ell^2v^2\left(v^2 + \frac{x}{\ell}\right).$$

Translating the xy-plane perpendicularly to itself, we obtain parabolic cylinders.

The equation of the half-plane passing through Ox is $v = 0$; the equation of its complement is $u = 0$. If we restrict v to being positive, we must allow u to be either positive or negative. The positive u will then define the positive y space; the negative u will give the negative y space. The equation u = constant will define only one half of a parabolic cylinder. The half-plane passing through Ox is the limiting parabolic cylinder with two sides to it: the upper side corresponding to positive u and the lower

side to negative u. Such coordinates are suitable for problems involving boundary conditions on the surface of a parabolic cylinder $v =$ constant. We could have restricted u to being positive and allowed all values to v; such coordinates would be suitable for problems involving boundary conditions on the surface of a cylinder $u =$ constant. The second frame of reference is just a mirror image of the first.

In terms of cartesian coordinates we have

$$u = \pm\sqrt{\frac{x + \sqrt{x^2 + y^2}}{2f}}, \qquad v = \sqrt{\frac{-x + \sqrt{x^2 + y^2}}{2f}};$$

the signs of u and y should be identical.

The fundamental metrical form is

$$ds^2 = 4f^2(u^2 + v^2)(du^2 + dv^2) + dz^2.$$

The expressions for the gradient, divergence, laplacian and curl are:

$$\text{grad}_u V = \frac{1}{2f\sqrt{u^2 + v^2}}\frac{\partial V}{\partial u}, \qquad \text{grad}_v V = \frac{1}{2f\sqrt{u^2 + v^2}}\frac{\partial V}{\partial v},$$

$$\text{grad}_z V = \frac{\partial V}{\partial z};$$

$$\text{div}\, F = \frac{1}{2f(u^2 + v^2)}\left[\frac{\partial}{\partial u}(\sqrt{u^2 + v^2}\, F_u) + \frac{\partial}{\partial v}(\sqrt{u^2 + v^2}\, F_v)\right] + \frac{\partial F_z}{\partial z};$$

$$\Delta V = \frac{1}{4f^2(u^2 + v^2)}\left(\frac{\partial^2 V}{\partial u^2} + \frac{\partial^2 V}{\partial v^2}\right) + \frac{\partial^2 V}{\partial z^2};$$

$$\text{curl}_u F = \frac{1}{2f\sqrt{u^2 + v^2}}\frac{\partial F_z}{\partial v} - \frac{\partial F_v}{\partial z},$$

$$\text{curl}_v F = \frac{\partial F_u}{\partial z} - \frac{1}{2f\sqrt{u^2 + v^2}}\frac{\partial F_z}{\partial u},$$

$$\text{curl}_z F = \frac{1}{2f(u^2 + v^2)}\left[\frac{\partial}{\partial u}(\sqrt{u^2 + v^2}\, F_v) - \frac{\partial}{\partial v}(\sqrt{u^2 + v^2}\, F_u)\right].$$

Paraboloidal coordinates. The paraboloidal coordinate system is obtained by taking the horizontal axis in Figure 8.6 as the z-axis, and making it the axis of revolution. The defining equations are

$$z + i\rho = f(u + iv)^2, \qquad z = f(u^2 - v^2), \qquad \rho = 2fuv;$$

$$ds^2 = 4f^2[(u^2 + v^2)(du^2 + dv^2) + u^2v^2\, d\varphi^2].$$

The expressions for the gradient, divergence, laplacian and curl are:

$$\operatorname{grad}_u V = \frac{1}{2\ell\sqrt{u^2+v^2}}\frac{\partial V}{\partial u}, \qquad \operatorname{grad}_v V = \frac{1}{2\ell\sqrt{u^2+v^2}}\frac{\partial V}{\partial v},$$

$$\operatorname{grad}_\varphi V = \frac{1}{2\ell uv}\frac{\partial V}{\partial \varphi};$$

$$\operatorname{div} F = \frac{1}{2\ell uv(u^2+v^2)}\left[\frac{\partial}{\partial u}(uv\sqrt{u^2+v^2}\,F_u) + \frac{\partial}{\partial v}(uv\sqrt{u^2+v^2}\,F_v)\right] + \frac{1}{2\ell uv}\frac{\partial F_\varphi}{\partial \varphi};$$

$$\Delta V = \frac{1}{4\ell^2 uv(u^2+v^2)}\left[v\frac{\partial}{\partial u}\left(u\frac{\partial V}{\partial u}\right) + u\frac{\partial}{\partial v}\left(v\frac{\partial V}{\partial v}\right)\right] + \frac{1}{4\ell^2u^2v^2}\frac{\partial^2 V}{\partial \varphi^2};$$

$$\operatorname{curl}_u F = \frac{1}{2\ell v\sqrt{u^2+v^2}}\frac{\partial}{\partial v}(vF_\varphi) - \frac{1}{2\ell uv}\frac{\partial F_v}{\partial \varphi},$$

$$\operatorname{curl}_v F = \frac{1}{2\ell uv}\frac{\partial F_u}{\partial \varphi} - \frac{1}{2\ell u\sqrt{u^2+v^2}}\frac{\partial}{\partial u}(uF_\varphi),$$

$$\operatorname{curl}_\varphi F = \frac{1}{2\ell(u^2+v^2)}\left[\frac{\partial}{\partial u}(\sqrt{u^2+v^2}\,F_v) - \frac{\partial}{\partial v}(\sqrt{u^2+v^2}\,F_u)\right].$$

CHAPTER IX

Exponential Functions

1. *Definitions*

In a broad sense, the name " exponential function " is applied to such functions as

$$y = 2^x, \qquad y = 10^x, \qquad y = a^x; \tag{1}$$

in a restricted sense it is applied to a particular function in which the " base " a is $e = 2.71828 \cdots$. In elementary arithmetic and algebra the exponent x is at first introduced as a shorthand notation for the number of times the base a is multiplied by itself. This definition requires x to be a positive integer. Later another definition is given for negative integral exponents, still another for fractional exponents, still another for imaginary exponents. These definitions are so formulated that the range of application of certain algebraic equations is extended. Thus if m and n are positive integers, then the definition of the exponent, as the number of times the base is repeated as a factor, leads at once to the following equations

$$a^m a^n = a^{m+n}, \qquad a^m/a^n = a^{m-n}. \tag{2}$$

Now, even for positive integral values of m and n, the second equation does not always have a meaning. If $m < n$, the exponent on the right side of the equation is negative and the base cannot possibly be repeated as a factor a negative number of times. On the other hand, the left side of the equation has a meaning for all positive integral values of m and n and thus can be used *to define* the meaning of the right side. Similarly, the right side of the first equation in (2) has a meaning when $m = n = \frac{1}{2}$ and thus can be used for defining the meaning of $a^{1/2}$ as the square root of a. Here we encounter our first difficulty: when x is a fraction, a^x becomes multiple valued and the number of values depends on x. If $x = \frac{1}{2}$, the function has two values; if $x = \frac{1}{3}$, the number of values is three; if $x = 0.01$, the number of values becomes one hundred.

Further difficulties arise when x is irrational. It is easy enough to define $10^{\sqrt{2}}$ as the limit of a sequence,

$$10^{1.4}, \qquad 10^{1.41}, \qquad 10^{1.414}, \cdots$$

provided we agree to take only the positive values of the roots; but if we try to apply the same definition to $(-10)^{\sqrt{2}}$, we find that the first approxi-

mation, $(-10)^{1.4} = (-10)^{7/5}$, is a negative real number, while the next, $(-10)^{1.41}$, is imaginary.

The best way to overcome these difficulties is to start anew and try to formulate a definition which will apply unambiguously and universally, first to all values of x and, then, to all values of a. Several such definitions have been found; perhaps the simplest is the one which defines exponential functions as *functions whose relative rates of change are constant*, that is, as *solutions of the following differential equation*

$$\frac{dw}{dz} = kw. \tag{3}$$

A particular exponential function whose relative rate of change k is unity, and whose value at $z = 0$ is also unity, is denoted by $\exp z$; that is $\exp z$ *is defined* by

$$\frac{d}{dz}\exp z = \exp z, \qquad \exp 0 = 1. \tag{4}$$

These definitions involve a radical change in the point of view. In elementary calculus we learn how to differentiate *known* functions; we have now introduced a new idea: a possibility of giving a rule for differentiation as a part of the definition of a function. It is important to become accustomed to this idea, since many other functions, such as Bessel and Legendre functions, are defined as solutions of certain differential equations. We should become used to the idea that differential equations may be regarded as a source of information about their solutions.

For example, equation (4) gives a rule for obtaining the approximate value of the increment of $\exp z$,

$$\Delta(\exp z) = \exp(z + \Delta z) - \exp z \simeq (\exp z)\,\Delta z. \tag{5}$$

Hence,

$$\exp(z + \Delta z) \simeq (1 + \Delta z)\exp z. \tag{6}$$

Starting with $z = 0$ where $\exp z$ is given, we obtain successively

$$\begin{aligned}
\exp(\Delta z) &\simeq (1 + \Delta z)\\
\exp(\Delta z + \Delta z) &\simeq (1 + \Delta z)\exp(\Delta z) = (1 + \Delta z)^2\\
\exp(2\Delta z + \Delta z) &\simeq (1 + \Delta z)\exp(2\Delta z) = (1 + \Delta z)^3\\
&\cdots\cdots\cdots\cdots\cdots\\
\exp(n\Delta z) &\simeq (1 + \Delta z)^n.
\end{aligned} \tag{7}$$

Let $n\Delta z = z$; then

$$\exp z \simeq \left(1 + \frac{z}{n}\right)^n, \qquad \exp z = \lim_{n\to\infty}\left(1 + \frac{z}{n}\right)^n. \tag{8}$$

From these equations we can obtain $\exp z$ as accurately as we wish; but the convergence is so extremely slow that this method is impractical.

A better expression is obtained as follows. By repeated differentiation of (4) we obtain any derivative of $\exp z$

$$\frac{d^2}{dz^2}\exp z = \frac{d}{dz}\exp z = \exp z, \qquad \cdots \frac{d^n}{dz^n}\exp z = \exp z. \tag{9}$$

Thus when $z = 0$, $\exp z$ and all its derivatives are equal to unity and by Maclaurin's theorem we obtain

$$\exp z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^n}{n!} + \cdots. \tag{10}$$

This series converges in the entire complex plane.

It should be noted that so far there are no indications whatsoever that the independent variable z in $\exp z$ is somehow related to the exponent in such expressions as 2^z, 10^z, etc., even when z is a positive integer. We might just as well have used some other symbol in place of "exp."

From the definition it follows that $w = \exp z$ is a solution of

$$\frac{dw}{dz} = w. \tag{11}$$

What is the general solution? The equation gives the derivative in terms of the function and by successive differentiation we may express all other derivatives in terms of w; but w itself is not given. If we assign an initial value $w = A$ for $z = 0$, then w will be determined for all values of z by Maclaurin's series,

$$\begin{aligned} w(z) &= A + Az + \frac{Az^2}{2!} + \frac{Az^3}{3!} + \cdots \\ &= A \exp z. \end{aligned} \tag{12}$$

Dividing (3) by k, and recalling that $d(kz) = kdz$, we find that the general solution of (3) is $A \exp (kz)$.

Problems

1. Calculate $\exp 1$ by (10), and then try using (8). *Ans.* $2.71828 \cdots$.
2. Calculate $\exp i$ by (10). *Ans.* $0.54030 + i0.84147$.
3. Calculate $\exp (0.1-0.2i)$. *Ans.* $1.08314 - i0.21956$.

2. *The addition theorem*

Further simplifications in the calculation of $\exp z$ may be made with the aid of the following *addition formula*

$$\exp (z_1 + z_2) = (\exp z_1)(\exp z_2). \tag{13}$$

In order to prove this we note that since $d(z+z_1) = dz$, $w = \exp(z+z_1)$ is a solution of (11). The value at $z = 0$ is $\exp z_1$; hence by (12) $w = (\exp z_1)(\exp z)$ and the theorem has been proved.

Another proof may be obtained if we replace all terms in (13) by the corresponding power series and show that the equation becomes an identity.

If z is large, the power series (10) is not very convenient for the evaluation of the function; but we can always choose $z = z_1 + z_2$, where z_1 and z_2 are smaller than z, and then use the addition formula (13).

For example, it is relatively easy to evaluate exp 1. Let us denote this value by the symbol "e"; then

$$e = 1 + 1 + \tfrac{1}{2} + \tfrac{1}{6} + \tfrac{1}{24} + \tfrac{1}{120} + \cdots = 2.71828\cdots. \tag{14}$$

From the addition formula we now obtain

$$\exp 2 = (\exp 1)(\exp 1) = e^2,$$
$$\exp 3 = (\exp 2)(\exp 1) = e^3,$$
$$\cdots\cdots\cdots\cdots\cdots$$

Thus for any positive integral value of z, we have

$$\exp z = e^z. \tag{15}$$

In every other case we simply define e^z as $\exp z$, whose value can always be uniquely obtained from (10). In other words, $\exp z$ and e^z are just two different notations for the function originally defined by (4) and then by (10).

Letting $z_1 = z$ and $z_2 = -z$ in (13), we have, in the new notation,

$$e^z e^{-z} = 1, \qquad e^{-z} = 1/e^z. \tag{16}$$

Likewise, we obtain

$$e^{x+iy} = e^x e^{iy}. \tag{17}$$

This formula is very convenient since it is easier to use the power series (10) when z is either real or pure imaginary, than when z is complex. In fact,

$$e^{iy} = \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \cdots\right) + i\left(y - \frac{y^3}{3!} + \frac{y^5}{5!} - \cdots\right). \tag{18}$$

The real and imaginary parts may now be recognized as power series for $\cos y$ and $\sin y$; thus

$$e^{iy} = \cos y + i \sin y. \tag{19}$$

However, we can obtain this expression directly from the formula (1–62) for differentiation of the unit complex number; (1–62) is a special case of (3) when $k = i$.

Problems

1. Show that $\exp\left(z \pm \frac{i\pi}{2}\right) = \pm i \exp z$.
2. Show that $\exp(z \pm i\pi) = -\exp z$.
3. Show that $\exp(z \pm 2\pi i) = \exp z$.
4. Show that $e^{2\pi i} = 1$.
5. Show that $\exp\left(z + \frac{i\pi}{4}\right) = \frac{1+i}{\sqrt{2}} \exp z$.

3. *Geometric interpretation of exponential functions of a complex variable*

It is possible to form a good picture of the behavior of the exponential function $w = \exp z = \exp(x + iy)$ from rather simple geometric considerations. Let us assume that x and y are proportional to time t

$$x = \xi t, \qquad y = \omega t,$$

$$z = (\xi + i\omega)t, \tag{20}$$

and see what happens to w when t receives an infinitesimal increment Δt. From the definition of $\exp z$ we have

$$\Delta w = w\Delta z = w\Delta x + iw\Delta y$$

$$= (\xi\Delta t)w + (i\omega\Delta t)w. \tag{21}$$

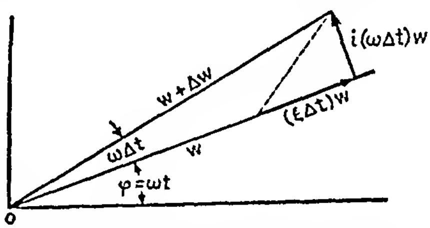

Fig. 9.1. If the relative increment of a complex variable w is $(\xi + i\omega)\,\Delta t$ where Δt is an infinitesimal interval of time, then w will revolve with the angular velocity ω and grow in magnitude at the relative rate ξ.

Thus Δw consists of two increments: one, $(\xi\Delta t)w$, in the direction of w; the other, $(i\omega\Delta t)w$, at right angles to w, Figure 9.1. Both increments are proportional to the magnitude of w. The first increment stretches w in the ratio $(1 + \xi\Delta t)/1$; the second rotates w through the angle $\omega\Delta t$. The stretching in the ratio $\sqrt{1 + (\omega\Delta t)^2}/1$ from the second increment is an infinitesimal of the second order. From all this we conclude that as t increases, w will turn counterclockwise with the angular velocity ω and its magnitude ρ will grow at the relative rate ξ, thus describing a spiral, Figure 9.2. If ξ is negative, ρ will decrease. The angle ψ, between w and Δw, may be found from Figure 9.1; thus $\tan\psi = \omega/\xi$. Hence the radii drawn from the origin to points on a particular spiral make equal angles with the spiral; such a spiral is called *equi-angular* or *logarithmic*.

4. Exp z *as the upper limit of an integral*

By definition, $w = \exp z$ is that particular solution of (11) for which $w = 1$ when $z = 0$. Solving (11) for dz,

$$dz = \frac{dw}{w}, \tag{22}$$

and integrating, we obtain

$$z = \int_1^w \frac{dw}{w}.$$

The lower limit of the integral has been made equal to unity in order to make z equal to zero when w is unity. The variable w in the integrand is a dummy variable; it is best to replace it by some other letter in order to stress that the value of the integral really depends on the upper limit

$$z = \int_1^w \frac{dt}{t}. \tag{23}$$

The upper limit in this integral is the exponential function of the value of the integral itself. When w is real, z may be regarded as the area under the

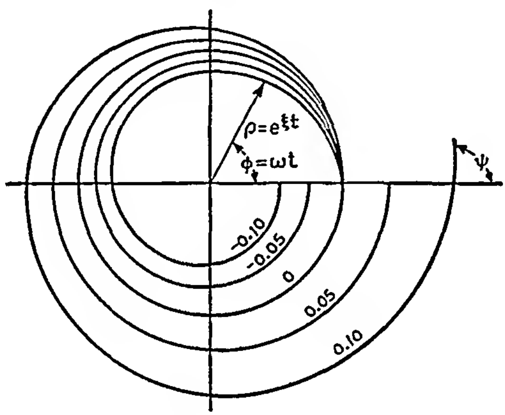

FIG. 9.2. Geometric representation of $w = e^{(\xi+\omega i)t}$ for several values of ξ.

hyperbola $s = 1/t$, between the ordinates corresponding to $t = 1$ and $t = w$. In this case, the second abscissa is the exponential function of the area.

The addition formula (13) may be obtained from (23) as follows. Let

$$z_1 = \int_1^{w_1} \frac{dt}{t}, \qquad z_2 = \int_1^{w_2} \frac{dr}{r}, \tag{24}$$

and add

$$z_1 + z_2 = \int_1^{w_1} \frac{dt}{t} + \int_1^{w_2} \frac{dr}{r}.$$

In the second integral substitute a new dummy variable t, defined by $t = w_1 r$,

$$z_1 + z_2 = \int_1^{w_1} \frac{dt}{t} + \int_{w_1}^{w_1 w_2} \frac{dt}{t} = \int_1^{w_1 w_2} \frac{dt}{t}. \tag{25}$$

Thus $w_1 w_2$ is $\exp(z_1 + z_2)$ as well as the product $(\exp z_1)(\exp z_2)$.

Equation (23) gives the value of w in terms of z only indirectly; in making a table we should have to assign a set of values to w and then evaluate by numerical integration the corresponding values of z. Regarded as a function of w, z is the inverse exponential function, $z = \exp^{-1} w$; this function is called the *logarithmic function*, $z = \log w$. In this notation equation (25) becomes $\log w_1 + \log w_2 = \log (w_1 w_2)$.

5. *Hyperbolic functions*

Hyperbolic functions may be defined in terms of exponential functions:

$$\cosh z = \tfrac{1}{2}(e^z + e^{-z}), \qquad \sinh z = \tfrac{1}{2}(e^z - e^{-z});$$

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{1 - e^{-2z}}{1 + e^{-2z}},$$

$$\coth z = \frac{\cosh z}{\sinh z} = \frac{e^z + e^{-z}}{e^z - e^{-z}} = \frac{1 + e^{-2z}}{1 - e^{-2z}}, \tag{26}$$

$$\operatorname{sech} z = \frac{1}{\cosh z}, \qquad \operatorname{csch} z = \frac{1}{\sinh z}.$$

Adding and subtracting the equations in the first line, we have

$$e^z = \cosh z + \sinh z,$$

$$e^{-z} = \cosh z - \sinh z. \tag{27}$$

Multiplying, we obtain

$$\cosh^2 z - \sinh^2 z = 1. \tag{28}$$

The resemblance between this equation and the equation

$$x^2 - y^2 = 1 \tag{29}$$

of an equilateral hyperbola, Figure 9.3, accounts for the name "hyperbolic functions." The added resemblance to $x^2 + y^2 = 1$ and $\cos^2\varphi + \sin^2\varphi = 1$ is responsible for the following more specific names given to the functions defined by (26): hyperbolic cosine, sine, tangent, cotangent, secant and cosecant.

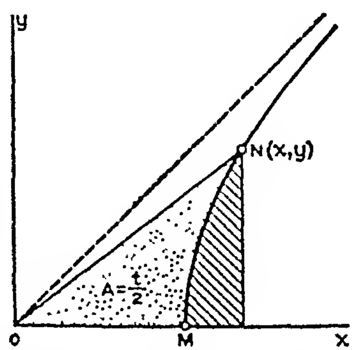

FIG. 9.3. If t is twice the dotted area MON where MN is an arc of the rectangular hyperbola $x^2 - y^2 = 1$, then the coordinates of N are $x = \cosh t$ and $y = \sinh t$.

If t is real, then

$$x = \cosh t, \qquad y = \sinh t \tag{30}$$

are parametric equations of the hyperbola (29). These equations are similar to the parametric equations of the circle of unit radius, Figure 9.4,

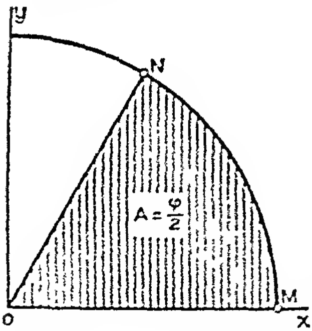

FIG. 9.4. If φ is twice the shaded area MON where MN is an arc of the circle $x^2 + y^2 = 1$, then the coordinates of N are $x = \cos\varphi$ and $y = \sin\varphi$.

$$x = \cos\varphi, \qquad y = \sin\varphi; \qquad x^2 + y^2 = 1. \tag{31}$$

Normally, the parameter φ is thought of as the length of the arc MN; however, in the case of the hyperbola the corresponding arc MN is not equal to t. The analogy holds only if we interpret φ as $2A$, where A is the area of the circular sector OMN; then t is also $2A$, where A is the area of a similar hyperbolic sector.

This geometric interpretation of hyperbolic functions is interesting, but it holds only for *real* values of the independent variable, while in practical applications the variable is frequently complex. The definitions (26) are really the best source of information about the hyperbolic functions.

Problems

1. Show that $\cosh 2z = \cosh^2 z + \sinh^2 z$ and $\sinh 2z = 2 \sinh z \cosh z$. *Hint:* Square and add the equations in the first line of (26); then take their product.

2. Show that $\dfrac{d \cosh z}{dz} = \sinh z$ and $\dfrac{d \sinh z}{dz} = \cosh z$.

3. Show that $\dfrac{d \tanh z}{dz} = \operatorname{sech}^2 z$. Show this in two ways, first using the results of the preceding problem, and then without using them.

4. Express $\cosh 3z$ and $\sinh 3z$ as polynomials in $\cosh z$ and $\sinh z$.

5. Express $\cosh^3 z$ and $\sinh^3 z$ in terms of the hyperbolic cosines and sines of multiples of z. *Hint:* Use (27).

6. Show that

$$\cosh z = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \frac{z^6}{6!} + \cdots$$

$$\sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \frac{z^7}{7!} + \cdots.$$

7. Show that

$$\cosh (z_1 + z_2) = \cosh z_1 \cosh z_2 + \sinh z_1 \sinh z_2,$$

$$\sinh (z_1 + z_2) = \sinh z_1 \cosh z_2 + \cosh z_1 \sinh z_2.$$

8. Show that $2\pi i$ is the smallest period of $\cosh z$ and $\sinh z$; and that $i\pi$ is the smallest period of $\tanh z$ and $\coth z$. What are the other periods?

9. Show that $\cosh (z + i\pi) = -\cosh z, \quad \sinh (z + i\pi) = -\sinh z,$

$$\cosh\left(z + \frac{i\pi}{2}\right) = i \sinh z, \quad \sinh\left(z + \frac{i\pi}{2}\right) = i \cosh z.$$

10. Find the values of $\cosh (1 + i)$, $\quad \sinh (1 + i)$, $\quad \tanh (0.5 + 2i)$.

11. Show that if y is small, then

$$\cosh (x + iy) \simeq \cosh x + iy \sinh x,$$

$$\sinh (x + iy) \simeq \sinh x + iy \cosh x.$$

6. *Circular functions*

The usual geometric definitions of the circular functions apply only to real values of the independent variable. A clue to the extension of these definitions to the entire complex plane may be found in equation (19). If we take the conjugate of this equation and then express $\cos y$ and $\sin y$ in terms of exponential functions, we obtain equations which can serve as definitions of circular functions for unrestricted values of y. Thus, using z in place of y, we have

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}), \qquad \sin z = \frac{1}{2i}(e^{iz} - e^{-iz}). \tag{32}$$

There are many functions which are defined for all z and which coincide with $\cos z$ and $\sin z$ for real z; but only one set is analytic. The derivative of an analytic function $f(z)$ is independent of the direction of approach to z; hence, in obtaining the derivatives of $\cos z$ at $z = 0$, we can approach the origin along the real axis where the cosine function is defined from geometric considerations; there

$$\begin{aligned} &\cos 0 = 1, \qquad \frac{d}{dz}\cos z = -\sin z = 0, \\ &\frac{d^2}{dz^2}\cos z = -\cos z = -1, \cdots. \end{aligned} \tag{33}$$

By Maclaurin's theorem we have

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots + (-)^n \frac{z^{2n}}{(2n)!} + \cdots. \tag{34}$$

Similarly,

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots + (-)^n \frac{z^{2n+1}}{(2n+1)!} + \cdots. \tag{35}$$

Comparing these series with those of Problem 6 in the preceding section, we have

$$\cos z = \cosh iz, \qquad i \sin z = \sinh iz. \tag{36}$$

Then from (26) we again find (32).

Problems

1. Show that

$$\cos(x + iy) = \cos x \cosh y - i \sin x \sinh y,$$
$$\sin(x + iy) = \sin x \cosh y + i \cos x \sinh y.$$

2. Show that, if y is small,

$$\cos(x + iy) \simeq \cos x - iy \sin x,$$
$$\sin(x + iy) \simeq \sin x + iy \cos x.$$

7. *Logarithmic and inverse hyperbolic functions*

In Section 4 we have defined the logarithmic function $z = \log w$ as the inverse of $w = \exp z$ and derived the addition theorem for it. This theorem can also be obtained from (13) by introducing $w_1 = \exp z_1$ and $w_2 = \exp z_2$; thus

$$\exp(z_1 + z_2) = w_1 w_2, \qquad z_1 + z_2 = \log(w_1 w_2),$$
$$\log w_1 + \log w_2 = \log(w_1 w_2). \tag{37}$$

From Problem 4 of Section 2 we have

$$\log 1 = 2n\pi i, \qquad n = 0, \pm 1, \pm 2, \cdots; \tag{38}$$

this shows that the logarithmic function possesses infinitely many branches.

Since $2\pi i$ is a period of $w = \exp z$, it is always possible to locate $z = \log w$ in the complex z-plane in the strip of width 2π, parallel to the real axis, Figure 9.5; this value of z is the *principal value* $(\log w)_p$ of $\log w$. All other values of the logarithm may be expressed in terms of the principal value

$$\log w = (\log w)_p + 2n\pi i. \tag{39}$$

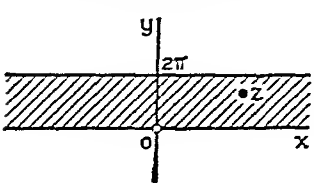

FIG. 9.5. If $z = \log w$, then corresponding to any point in the w-plane we can find a single point z in the shaded strip of the z-plane.

The principal value of $\log 1$ is zero and the principal value, $\log \rho$, of any positive number ρ is real; hence for any complex number

$$\log(\rho e^{i\varphi}) = \log \rho + i\varphi + 2n\pi i. \tag{40}$$

The phase of the principal value is equal to or greater than zero but less than 2π. Sometimes, however, it may be convenient to place the principal values in the strip $-\pi < \varphi \leq \pi$.

Since $a = e^{\log a}$, a natural definition of a^z is

$$a^z = e^{z \log a}. \tag{41}$$

Inverse hyperbolic functions $\cosh^{-1} w$, $\sinh^{-1} w$, ... and inverse circular functions $\cos^{-1} w$, $\sin^{-1} w$, ... are also many-valued functions and can be expressed in terms of logarithmic functions. For example,

$$\begin{gathered} z = \cosh^{-1} w, \qquad w = \cosh z = \tfrac{1}{2}(e^z + e^{-z}), \\ e^{2z} - 2we^z + 1 = 0, \qquad e^z = w \pm \sqrt{w^2 - 1}, \\ z = \cosh^{-1} w = \log (w \pm \sqrt{w^2 - 1}). \end{gathered} \tag{42}$$

Similarly,

$$\cos^{-1} w = i \log (w \pm \sqrt{w^2 - 1}) = i \log (w \pm i\sqrt{1 - w^2}). \tag{43}$$

All inverse hyperbolic and circular functions can be expressed in logarithmic form.

Problems

1. Find $\log (1 + i)$.

Ans. $0.346 \cdots + \dfrac{i\pi}{4} + 2n\pi i.$

2. Find $\log (-1 + 2i)$.

Ans. $0.805 - i\,1.107 + 2n\pi i.$

3. Find the derivatives of $\cosh^{-1} z$, $\sinh^{-1} z$, $\tanh^{-1} z$.

Ans. $1/\sqrt{z^2 - 1}, \quad 1/\sqrt{1 + z^2}, \quad 1/(1 - z^2).$

4. Show that $\sinh^{-1} z = \log (z \pm \sqrt{z^2 + 1})$, $\tanh^{-1} z = \dfrac{1}{2} \log \dfrac{1 + z}{1 - z}$.

5. Obtain the following Fourier series

$$e^{\rho \cos \varphi} \cos (\rho \sin \varphi) = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} \cos n\varphi,$$

$$e^{\rho \cos \varphi} \sin (\rho \sin \varphi) = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} \sin n\varphi.$$

6. Show that $D_z (\log z) = 1/z$.

7. Derive the following series

$$\log (1 + z) = z - \tfrac{1}{2} z^2 + \tfrac{1}{3} z^3 - \tfrac{1}{4} z^4 + \cdots,$$

$$\log (1 + z) = \log z + z^{-1} - \tfrac{1}{2} z^{-2} + \tfrac{1}{3} z^{-3} - \tfrac{1}{4} z^{-4} + \cdots.$$

Note that the first series represents only that branch of $\log (1 + z)$ for which $\log 1 = 0$.

Show that the first series converges absolutely if $\rho = |z| < 1$ and relatively if $\rho = 1$

but $\varphi \neq (2n+1)\pi$. Similarly show that the second series converges absolutely if $\rho > 1$.

Hint: Note that the first series may be obtained from the geometric series by integration and that the remainder term may be expressed as an integral.

8. Derive the following Fourier series

$$\log (1 - 2\rho \cos \varphi + \rho^2) = -2 \sum_{n=1}^{\infty} \frac{1}{n} \rho^n \cos n\varphi,$$

$$\tan^{-1} \frac{\rho \sin \varphi}{1 - \rho \cos \varphi} = \sum_{n=1}^{\infty} \frac{1}{n} \rho^n \sin n\varphi.$$

In particular show that

$$\sum_{n=1}^{\infty} \frac{\cos n\varphi}{n} = \frac{1}{2} \log \frac{1}{2(1 - \cos \varphi)} = -\log [2 \sin (\varphi/2)], \qquad 0 < \varphi < 2\pi;$$

$$\sum_{n=1}^{\infty} \frac{\sin n\varphi}{n} = \frac{\pi - \varphi}{2}, \qquad 0 < \varphi < 2\pi.$$

9. Show that $\sum_{n=1}^{\infty} \frac{\cos n\varphi}{n^2} = \frac{\pi^2}{6} - \frac{\pi\varphi}{2} + \frac{\varphi^2}{4}$. Assume that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ is a known result.

10. Prove that

$$\cosh^{-1} x + \cosh^{-1} y = \cosh^{-1} [xy + \sqrt{(x^2 - 1)(y^2 - 1)}].$$

Hint: Let $\cosh^{-1} x = u$, $\cosh^{-1} y = v$, $\cosh^{-1} x + \cosh^{-1} y = w$; then $\cosh w = \cosh (u + v)$, etc.

11. Prove that

$$\tanh^{-1} x + \tanh^{-1} y = \tanh^{-1} \frac{x + y}{1 + xy}.$$

8. *Exponential functions of time and distance*

We are now in a position to generalize the method explained in Section 1.10. There the analysis of steady state harmonic oscillations, which normally requires the solution of certain differential equations, is reduced to the solution of simple algebraic equations. The success of the method depends, aside from the linearity of the equations, on the fact that the derivatives of the time factor $\mathcal{T} = \exp (i\omega t)$ are proportional to $\mathcal{T}$ and that, consequently, $\mathcal{T}$ cancels out. This is still true if $\mathcal{T}$ is of more general form

$$\mathcal{T} = e^{pt}, \qquad d\mathcal{T}/dt = p\mathcal{T}. \tag{44}$$

The complex number

$$p = \xi + i\omega \tag{45}$$

is called the *oscillation constant;* the real part ξ is the *growth constant.*

In the case of the oscillations of mass M attached to a spring, (1–77) now becomes

$$y = \frac{F}{(S + p^2M) + pR}. \tag{46}$$

Similarly for an electric network consisting of a resistance R, inductance L and capacitance C, Figure 9.6, the complex current I is given in terms of the complex voltage V by

$$I = \frac{V}{R + pL + (1/pC)}. \tag{47}$$

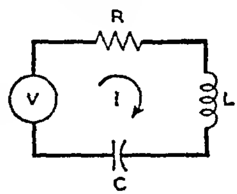

FIG. 9.6. A one-mesh electric circuit.

More generally, for any linear network,

$$I = \frac{V}{Z(p)} = Y(p)V, \tag{48}$$

where the *impedance function* $Z(p)$ and *admittance function* $Y(p)$ are analytic functions of the complex variable p. The properties of these functions determine the behavior of the network.

The *transient* or *natural oscillations* of the system are included in this analysis. If the force applied to the mass M is zero, and if $\bar{y} =$ re $[y \exp (pt)]$, (1–73) becomes

$$(Mp^2 + Rp + S)y = 0. \tag{49}$$

There is one trivial solution, $y = 0$, which corresponds to a system remaining at rest. For the other solution we have

$$Mp^2 + Rp + S = 0, \qquad p = -\frac{R}{2M} \pm \sqrt{\frac{R^2}{4M^2} - \frac{S}{M}}, \tag{50}$$

with y remaining arbitrary. This does not mean that there can be motion without *any* force applied to the mass; what it does mean is that motion can exist without a *continuously applied force*.

Natural oscillations are truly oscillatory only if $R < 2\sqrt{SM}$; then

$$p = -\frac{R}{2M} \pm i\sqrt{\frac{S}{M} - \frac{R^2}{4M^2}}; \qquad \xi = -\frac{R}{2M},$$
$$\omega = \sqrt{\frac{S}{M} - \frac{R^2}{4M^2}}. \tag{51}$$

The quantity $R/2M$ is the *damping constant;* ω is the *natural frequency* in radians per second. In order to obtain the instantaneous values of the

displacement we proceed as follows:

$$\begin{aligned}\tilde{y} &= \text{re}\,(ye^{pt}) = \text{re}\,(y_1e^{\xi+i\omega t} + y_2e^{\xi-i\omega t})\\ &= \text{re}\,(y_1e^{\xi}\cos\omega t + iy_1e^{\xi}\sin\omega t + y_2e^{\xi}\cos\omega t - iy_2e^{\xi}\sin\omega t)\\ &= Ae^{\xi}\cos\omega t + Be^{\xi}\sin\omega t.\end{aligned} \tag{52}$$

In passing to the last form we should remember that y_1 and y_2 are arbitrary; hence they can be chosen to make $A = y_1 + y_2$ and $B = i(y_1 - y_2)$ real, by letting $y_1 = \frac{1}{2}(A - iB)$ and $y_2 = \frac{1}{2}(A + iB)$.

Exponential functions occur also in the theory of waves. Consider the following function of distance and time

$$\psi = A\exp(\omega t - \gamma x), \qquad A = a\exp(i\vartheta), \tag{53}$$

where the complex number

$$\gamma = \alpha + i\beta \tag{54}$$

is called the *propagation constant* for reasons that will presently become obvious. Such functions are called *wave functions*. The phase Φ of ψ is

$$\Phi = \omega t - \beta x + \vartheta. \tag{55}$$

It is a linear function of time at any given place and a linear function of distance at any given instant. At any particular place we should observe oscillations of ψ; at any instant the " frozen " profile of ψ is a sinusoid of gradually changing amplitude.

The phase appears stationary if

$$d\Phi = \omega\,dt - \beta\,dx = 0, \qquad dx/dt = \omega/\beta; \tag{56}$$

that is, if the observer is moving with the velocity

$$v = \omega/\beta. \tag{57}$$

This is the *phase velocity*.

Two phases differing by 2π are indistinguishable. The period of oscillations, T, is the time required for the phase to increase by 2π; similarly, the *wavelength*, λ, is the change in distance which corresponds to an increase in phase of 2π. The analogy between the angular frequency ω and the period T on one side, and the *phase constant* β and the wavelength λ on the other, is reflected in the following equations

$$\begin{aligned}\omega T &= 2\pi, \qquad T = 2\pi/\omega, \qquad \omega = 2\pi/T;\\ \beta\lambda &= 2\pi, \qquad \lambda = 2\pi/\beta, \qquad \beta = 2\pi/\lambda.\end{aligned} \tag{58}$$

The *wave number*, $\gamma = 1/\lambda$, representing the number of full waves per unit length, is analogous to the frequency, $f = 1/T$, representing the number of full cycles per unit time.

In natural waves the real part α of the propagation constant is negative and for this reason its absolute value is called the *attenuation constant.* It is analogous to the damping constant. In equation (53) the wave function represents sustained oscillations. The more general form is

$$\psi = A \exp(pt - \gamma x), \tag{59}$$

where p and γ are both complex.

Problems

1. Find the "steady state" solutions of the following equations:

a) $y' + 2y = 4e^{3t}$, b) $2y'' - 3y' + y = 5e^{2t}$,

c) $y' + y = e^t \cos 5t$, d) $y''' - 2y = 2e^t \sin t$,

e) $y'' + y = 3e^{-t} + 2e^t + e^{2t} \cos t$.

Note: In case (c) the right side may be regarded as re $[\exp(1 + 5i)t]$ and in (d) as re $[-2i \exp(1 + i)t]$ or as im $[2 \exp(1 + i)t]$; in case (e) solutions may be found for each righthand side term and then added.

2. Find the "transient" solutions of the above equations.

9. *A collection of formulas*

The following collection of formulas is presented for reference. With profit, the student may regard them as review problems.

(1) $e^z = 1 + z + \dfrac{z^2}{2!} + \dfrac{z^3}{3!} + \cdots + \dfrac{z^n}{n!} + \cdots.$

(2) $\cosh z = 1 + \dfrac{z^2}{2!} + \dfrac{z^4}{4!} + \cdots + \dfrac{z^{2n}}{(2n)!} + \cdots.$

(3) $\sinh z = z + \dfrac{z^3}{3!} + \dfrac{z^5}{5!} + \cdots + \dfrac{z^{2n+1}}{(2n+1)!} + \cdots.$

(4) $\tanh z = z - \frac{1}{3}z^3 + \frac{2}{15}z^5 - \frac{17}{315}z^7 + \cdots, \quad |z| < \frac{1}{2}\pi.$

(5) $\coth z = z^{-1} + \frac{1}{3}z - \frac{1}{45}z^3 + \frac{2}{945}z^5 - \cdots, \quad |z| < \pi.$

(6) $\cos z = 1 - \dfrac{z^2}{2!} + \dfrac{z^4}{4!} - \cdots + (-)^n \dfrac{z^{2n}}{(2n)!} + \cdots.$

(7) $\sin z = z - \dfrac{z^3}{3!} + \dfrac{z^5}{5!} - \cdots + (-)^n \dfrac{z^{2n+1}}{(2n+1)!} + \cdots.$

(8) $\tan z = z + \frac{1}{3}z^3 + \frac{2}{15}z^5 + \frac{17}{315}z^7 + \cdots, \quad |z| < \frac{1}{2}\pi.$

(9) $\cot z = z^{-1} - \frac{1}{3}z - \frac{1}{45}z^3 - \frac{2}{945}z^5 - \frac{1}{4725}z^7 - \cdots, \quad |z| < \pi.$

(10) $e^{u+v} = e^u e^v, \quad e^{x+iy} = e^x (\cos y + i \sin y).$

(11) $\cosh (u \pm v) = \cosh u \cosh v \pm \sinh u \sinh v.$

(12) $\sinh (u \pm v) = \sinh u \cosh v \pm \cosh u \sinh v.$

(13) $\tanh (u \pm v) = \dfrac{\tanh u \pm \tanh v}{1 \pm \tanh u \tanh v}.$

(14) $\cos (u \pm v) = \cos u \cos v \mp \sin u \sin v.$

(15) $\sin (u \pm v) = \sin u \cos v \pm \cos u \sin v.$

(16) $\tan (u \pm v) = \dfrac{\tan u \pm \tan v}{1 \mp \tan u \tan v}.$

(17) $\cosh (x + iy) = \cosh x \cos y + i \sinh x \sin y.$

(18) $\cos (x + iy) = \cos x \cosh y - i \sin x \sinh y.$

(19) $\sinh (x + iy) = \sinh x \cos y + i \cosh x \sin y.$

(20) $\sin (x + iy) = \sin x \cosh y + i \cos x \sinh y.$

(21) $\cosh iz = \cos z, \quad \sinh iz = i \sin z, \quad \tanh iz = i \tan z.$

(22) $\cos iz = \cosh z, \quad \sin iz = i \sinh z, \quad \tan iz = i \tanh z.$

(23) $\cosh \left(z \pm \dfrac{i\pi}{2}\right) = \pm i \sinh z, \quad \cosh (z \pm i\pi) = -\cosh z.$

(24) $\sinh \left(z \pm \dfrac{i\pi}{2}\right) = \pm i \cosh z, \quad \sinh (z \pm i\pi) = -\sinh z.$

(25) $\cos \left(z \pm \dfrac{\pi}{2}\right) = \mp \sin z, \quad \cos (z \pm \pi) = - \cos z.$

(26) $\sin \left(z \pm \dfrac{\pi}{2}\right) = \pm \cos z, \quad \sin (z \pm \pi) = - \sin z.$

(27) $\tanh \left(z \pm \dfrac{i\pi}{2}\right) = \coth z, \quad \tanh (z \pm i\pi) = \tanh z.$

(28) $\tan \left(z \pm \dfrac{\pi}{2}\right) = - \cot z, \quad \tan (z \pm \pi) = \tan z.$

(29) $\log u + \log v = \log (uv).$

(30) $\log (\rho e^{i\varphi}) = \log \rho + i\varphi + 2n\pi i, \quad n = 0, \quad \pm 1, \quad \pm 2, \cdots.$

(31) $\log (1 +)z = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots.$

(32) $\cosh^{-1} z = \log(z \pm \sqrt{z^2 - 1}), \quad \sinh^{-1} z = \log(z \pm \sqrt{z^2 + 1}).$

(33) $\tanh^{-1} z = \frac{1}{2} \log \frac{1 + z}{1 - z}, \quad \coth^{-1} z = \frac{1}{2} \log \frac{z + 1}{z - 1}.$

(34) $\cos^{-1} z = i \log(z \pm i\sqrt{1 - z^2}), \quad \sin^{-1} z = i \log(-iz \pm \sqrt{1 - z^2}).$

(35) $\tan^{-1} z = \frac{1}{2} i \log \frac{1 - iz}{1 + iz}, \quad \cot^{-1} z = \frac{1}{2} i \log \frac{z - i}{z + i}.$

(36) $\tanh z = \sum_{n=1}^{\infty} \frac{8z}{(2n - 1)^2\pi^2 + 4z^2}, \quad \coth z = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{n^2\pi^2 + z^2}.$

(37) $\tan z = \sum_{n=1}^{\infty} \frac{8z}{(2n - 1)^2\pi^2 - 4z^2}, \quad \cot z = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2\pi^2}.$

(38) $\operatorname{sech} z = \sum_{n=1}^{\infty} \frac{(-)^{n-1}4(2n - 1)\pi}{(2n - 1)^2\pi^2 + 4z^2}, \quad \operatorname{csch} z = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{(-)^n 2z}{z^2 + n^2\pi^2}.$

(39) $\sec z = \sum_{n=1}^{\infty} \frac{(-)^n 4(2n - 1)\pi}{4z^2 - (2n - 1)^2\pi^2}, \quad \csc z = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{(-)^n 2z}{z^2 - n^2\pi^2}.$

CHAPTER X

Differential Equations of the First Order

Differential equations are classified into *ordinary* differential equations, containing ordinary derivatives, and *partial* differential equations, containing partial derivatives. They are subclassified according to their *order* and *degree:* the order of a differential equation is the order of the highest derivative; the degree of the equation is the highest degree of the highest order derivative. The equation is *linear* if the dependent variable and its derivatives occur in the first degree without any cross products; otherwise the equation is *non-linear.* The independent variable in a linear equation may occur in any degree or as the argument of a transcendental function. Every linear equation is automatically of the first degree; but not every equation of the first degree is linear. A linear equation is *homogeneous* if every term contains either the dependent variable or one of its derivatives.

This and the following chapters are devoted almost exclusively to linear equations because of their importance in applied mathematics and particularly because of the relative simplicity of their theory. Non-linear equations would take us too far afield. In this chapter we are concerned with first-order equations; in the next we shall treat second-order linear equations; and in the following chapter we shall deal briefly with equations of higher orders, mostly with the case in which the coefficients are constants.

1. *Linear equations*

The most general linear equation of the first order is

$$\frac{dy}{dx} = P(x)y + Q(x), \tag{1}$$

where P and Q are arbitrary functions.

Integral calculus is concerned with solutions of the special case in which $P(x) = 0$. The theory of exponential functions may be regarded — as in the preceding chapter — as the theory of the solutions of another special case, in which $Q(x) = 0$ and $P(x)$ is a constant. The next case to be considered is the homogeneous equation, in which $Q(x)$ vanishes but $P(x)$ is not necessarily constant.

2. *Solution of the homogeneous equation*

The homogeneous equation

$$\frac{dy}{dx} = P(x)y \tag{2}$$

can be immediately reduced to an equation with constant coefficients by changing the independent variable; thus we write

$$\psi = \int_a^x P(x)\,dx, \qquad \frac{d\psi}{dx} = P(x), \tag{3}$$

where a is a constant. Substituting in (2),

$$\frac{dy}{dx} = \frac{dy}{d\psi}\frac{d\psi}{dx} = P(x)\frac{dy}{d\psi}, \qquad \frac{dy}{d\psi} = y. \tag{4}$$

Hence the solution is

$$y(x) = A \exp \psi = A \exp \int_a^x P(x)\,dx. \tag{5}$$

If $x = a$, the integral vanishes and the exponential function reduces to unity; therefore $A = y(a)$ and

$$y(x) = y(a) \exp \int_a^x P(x)\,dx.$$

Problems

1. Solve $dy/dx = xy$, $y = 1$ if $x = 0$. *Ans.* $y(x) = \exp(\frac{1}{2}x^2)$.

2. Solve the above equation by computing successive derivatives and using Maclaurin's formula.

3. Solve the above equation by assuming at the start a power series $y = a_0 + a_1x + a_2x^2 + \cdots$. Substitute the series in the equation and equate the coefficients of the various powers of x.

3. *The nonhomogeneous equation with constant P*

The equation to be considered is

$$\frac{dy}{dx} = ky + Q(x). \tag{6}$$

The fundamental idea underlying the following method of solution is worth special attention, for it is invaluable in the solution of many physical problems. It can best be explained by first considering a concrete example, and then generalizing.

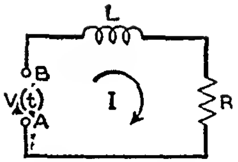

FIG. 10.1. An arbitrary voltage $V(t)$ applied to an electric circuit consisting of an inductance L in series with a resistance R.

Take an electric circuit consisting of a resistance R and inductance L in series, Figure 10.1; and let $V(t)$ be the impressed voltage. If $I(t)$ is the current, then

$$L\frac{dI}{dt} + RI = V(t), \qquad \frac{dI}{dt} = -\frac{R}{L}I + \frac{1}{L}V(t). \tag{7}$$

If the impressed voltage were zero *at all times*, the response would also be zero at all times. Suppose now that $V(t)$ is equal to zero at all times *except* in the interval $(\tau,\tau + \Delta\tau)$. Integrating (7) over this interval, we have

$$I(\tau + \Delta\tau) - I(\tau) = -\frac{R}{L}\int_{\tau}^{\tau+\Delta\tau} I(t)\,dt + \frac{1}{L}\int_{\tau}^{\tau+\Delta\tau} V(t)\,dt. \tag{8}$$

Let $V(t)$ increase indefinitely and $\Delta\tau$ approach zero in such a way that the *voltage impulse* remains unity

$$\int_{\tau}^{\tau+\Delta\tau} V(t)\,dt = 1. \tag{9}$$

If $I(t)$ remains finite, the first integral on the right of (8) approaches zero with $\Delta\tau$ and the equation becomes

$$I(\tau + 0) - I(\tau) = 1/L. \tag{10}$$

In other words at $t = \tau$ there is a sudden rise in current. Prior to this instant, and subsequently, $V(t)$ is zero, equation (7) is homogeneous, and its solution is of the form

$$I(t) = A\exp(-Rt/L). \tag{11}$$

For $t < \tau$, $I(t)$ vanishes and therefore A must vanish prior to this instant; at $t = \tau$ there should be an increment equal to $1/L$; hence,

$$I(t) = \frac{1}{L}\exp\frac{-R(t-\tau)}{L}, \quad t > \tau. \tag{12}$$

This solution is shown graphically in Figure 10.2.

The equation is linear and the response is proportional to the voltage impulse; if the latter is $V(\tau)\,d\tau$, instead of unity, (12) is multiplied by this factor. Considering the continuous action of $V(t)$ as the limit of a succession of impulses, we have

$$I(t) = \frac{1}{L}\int_{a}^{t} V(\tau)\exp\frac{-R(t-\tau)}{L}\,d\tau, \tag{13}$$

for the case in which $V(t)$ vanishes for $t < a$. Otherwise, $a = -\infty$.

Thus the electric circuit in Figure 10.1 " solves " the general equation (6):

$$y(x) = \int_a^x Q(\xi)e^{k(x-\xi)}\, d\xi = e^{kx} \int_a^x Q(\xi)e^{-k\xi}\, d\xi. \tag{14}$$

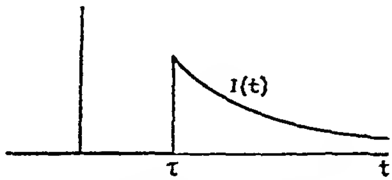

Fig. 10.2. The response of the circuit shown in Figure 10.1 when the applied voltage is an impulse at $t = \tau$.

To verify, differentiate this equation and substitute in (6):

$$\frac{dy}{dx} = ke^{kx} \int_a^x Q(\xi)e^{-k\xi}\, d\xi + e^{kx}\frac{d}{dx}\int_a^x Q(\xi)e^{-k\xi}\, d\xi$$

$$= ke^{kx} \int_a^x Q(\xi)e^{-k\xi}\, d\xi + e^{kx}Q(x)e^{-kx} = ky + Q(x). \tag{15}$$

Some readers may find it easier to think in terms of a mass M moving through a resisting medium with a velocity v under the influence of a force $F(t)$ rather than in terms of an electrical circuit. An impulsive force is a familiar conception in dynamics; that the impulse of force is equal to the increase in momentum is a familiar law; with suitable changes in wording the above discussion may be recast around the example from dynamics. In mathematics, the impulse function is frequently called the *δ-function*.

4. *The general nonhomogeneous equation*

The general case in which P and Q are both variable may also be solved with the aid of the impulse or delta function. We shall use this opportunity, however, to explain another method, *the method of variation of parameter*. The solution (5) of the homogeneous equation contains a " constant of integration " A; the question is, can we obtain the solution of the nonhomogeneous case by assuming A to be some function of x? Let us try. Differentiating $y(x)$ subject to this assumption,

$$\frac{dy}{dx} = \frac{dA}{dx} \exp \int_a^x P(x)\, dx + AP \exp \int_a^x P(x)\, dx, \tag{16}$$

and substituting in (1), we have

$$\frac{dA}{dx}\exp\int_a^x P(x)\,dx = Q(x), \qquad \frac{A}{dx} = Q(x)\exp\left[-\int_a^x P(x)\,dx\right], \tag{17}$$

$$A(x) = \int_a^x Q(x)\exp\left[-\int_a^x P(x)\,dx\right]dx + \text{constant}.$$

Thus the answer to our question is affirmative and the solution is

$$y(x) = e^{\int_a^x P(x)dx}\int_a^x Q(x)e^{-\int_a^x P(x)dx}\,dx + Ce^{\int_a^x P(x)dx}. \tag{18}$$

The first term represents a particular solution which vanishes at $x = a$: the second term is the *complementary solution* which assumes the value C at $x = a$.

This method of solution is not really very different from the one we used earlier; both methods are essentially *perturbation methods*. The solution of the homogeneous equation is perturbed to fit the nonhomogeneous case. In fact we can transform (18) to make it correspond to the form we would have obtained by the first method. To this end, we move the exponential factor under the sign of integration. In so doing we must distinguish between the "x" which occurs as one of the limits of integration and the "x" which is used as a dummy variable of integration. Denoting the latter by ξ or η, we rewrite (18) as follows:

$$y(x) = e^{\int_a^x P(\eta)d\eta}\int_a^x Q(\xi)e^{-\int_a^\xi P(\eta)d\eta}\,d\xi + Ce^{\int_a^x P(\eta)d\eta}. \tag{19}$$

Now we are free to move the exponential factor under the sign of integration; thus

$$y(x) = Ce^{\int_a^x P(\eta)d\eta} + \int_a^x Q(\xi)e^{\int_\xi^x P(\eta)d\eta}\,d\xi. \tag{20}$$

If $P(x) = k$, this equation reduces to (14), except for the first term which is not essential if a is taken as a parameter. In fact,

$$\int_a^x Q(\xi)e^{-k\xi}\,d\xi = \int_a^{a_1} Q(\xi)e^{-k\xi}\,d\xi + \int_{a_1}^x Q(\xi)e^{-k\xi}\,d\xi; \tag{21}$$

and if a_1 is kept fixed while a is varied, the first term in (21) becomes the variable parameter "C" and (14) assumes the form of (20).

The exponential factor in the integrand of (20) is that particular solution of the *homogeneous* equation which reduces to unity at $x = \xi$; thus equation (20) may be interpreted as the solution of the homogeneous equation which varies in magnitude by infinitesimal jumps equal to $Q(\xi)\,d\xi$.

Problems

1. Solve $dy/dx = xy + x$, $y = 1$ if $x = 0$. *Ans.* $y = 2\exp(\frac{1}{2}x^2) - 1$.

2. Show that more generally the solution of the above equation is $y(x) = [y(0) + 1]\exp(\frac{1}{2}x^2) - 1$.

3. Solve the above equation by the power series method.

4. Show that if $f(x)$ is any particular solution of (1), then the general solution is $y(x) = y_0(x) + f(x)$ where $y_0(x)$ is the general solution of (2). *Hint:* Substitute the suggested form in (1) and take into consideration the assumed property of $f(x)$.

This theorem may occasionally reduce the effort of solving a given equation. Thus it is evident by inspection that $y = -1$ is a particular solution of the equation in Problem 1; from there on we need consider only the homogeneous equation.

5. *Nonlinear equations — Picard's method*

A general nonlinear equation of the first order is of the form

$$F\left(\frac{dy}{dx}, y, x\right) = 0. \tag{22}$$

Let us assume that we can solve this equation for the first derivative

$$\frac{dy}{dx} = f(x,y). \tag{23}$$

Picard's method consists of integrating (23) term by term in the interval (a,x)

$$y(x) - y(a) = \int_a^x f(\xi,y)\,d\xi, \qquad y(x) = y(a) + \int_a^x f(\xi,y)\,d\xi. \tag{24}$$

Using $y(a)$ as the first approximation, we substitute it in the integral and obtain the second approximation. In the same way we obtain the third approximation from the second, etc.

If we wish to make a table of approximate values of $y(x)$, the procedure is not very different from that used for numerical integration in Chapter 6. Choosing a small interval h, we rewrite (24) as follows:

$$y(x) = y(a) + \int_a^{a+h} f(\xi,y)\,d\xi + \int_{a+h}^{a+2h} f(\xi,y)\,d\xi$$
$$+ \int_{a+2h}^{a+3h} f(\xi,y)\,d\xi + \cdots \int_{a+nh}^{x} f(\xi,y)\,d\xi. \tag{25}$$

To the extent to which $f(\xi,y)$ is constant in each interval, we have

$$y(x) = y(a) + hf[a,y(a)] + hf[a+h, y(a+h)] + hf[a+2h, y(a+2h)]$$
$$+ \cdots + (x - a - nh)f[a+nh, y(a+nh)]; \tag{26}$$

that is, each increment of y is obtained from the value of y at the end of the preceding interval. This method was first proposed by Cauchy; but it is seen to be closely related to Picard's method.

For instance, let

$$\frac{dy}{dx} = \sin(x + y), \qquad y = 0.5 \text{ if } x = 0. \tag{27}$$

Choosing $h = 0.1$, we have

$$\begin{aligned} y(0) &= 0.5, \qquad y(0.1) = 0.5 + 0.1 \sin 0.5 = 0.5479, \\ y(0.2) &= 0.5479 + 0.1 \sin 0.6479 = 0.6083. \end{aligned} \tag{28}$$

The results are summarized in the following table:

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$y(x)$	0.5	0.5479	0.6083	0.6806	0.7637	0.8555	0.9532	1.053	1.151	1.244	1.328

(29)

Equation (27) can be solved exactly in terms of known functions, and it is found that $y(1) = 1.344$. Considering that the interval of integration is rather large, our table is quite good. For greater accuracy the interval should be diminished. With the aid of a computing machine the task is not onerous; and there exist modern machines which perform the required operations automatically.

In some cases it is possible to obtain a solution of (24) as an infinite sequence by successive substitutions. Take, for instance,

$$\begin{aligned} \frac{dy}{dx} &= y^2, \qquad y = 1 \text{ if } x = 0; \\ y(x) &= 1 + \int_0^x [y(\xi)]^2 \, d\xi. \end{aligned} \tag{30}$$

The sequence of successive approximations is

$$\begin{aligned} y_0(x) &= 1, \qquad y_1(x) = 1 + \int_0^x d\xi = 1 + x, \\ y_2(x) &= 1 + \int_0^x (1 + \xi)^2 \, d\xi = 1 + \tfrac{1}{3}[(1 + x)^3 - 1], \\ y_2(x) &= \tfrac{2}{3} + \tfrac{1}{3}(1 + x)^3 = 1 + x + x^2 + \tfrac{1}{3}x^3, \\ y_3(x) &= 1 + \tfrac{1}{9}\int_0^x [4 + 4(1 + \xi)^3 + (1 + \xi)^6] \, d\xi \\ &= 1 + \tfrac{4}{9}x + \tfrac{1}{9}(1 + x)^4 - \tfrac{1}{9} + \tfrac{1}{63}(1 + x)^7 - \tfrac{1}{63} \\ &= 1 + x + x^2 + x^3 + \tfrac{2}{3}x^4 + \tfrac{1}{3}x^5 + \tfrac{1}{9}x^6 + \tfrac{1}{63}x^7. \end{aligned} \tag{31}$$

Again we can compare these approximations with the exact solution. Regarding x as the dependent variable, we obtain from (30)

$$dx = \frac{dy}{y^2}, \qquad x = \int_1^y \frac{dy}{y^2} = 1 - \frac{1}{y}, \qquad y = \frac{1}{1-x}. \tag{32}$$

Thus the first four terms of $y_3(x)$ coincide with the corresponding terms of the exact power series.

It should be noted, however, that unless the interval is small, Picard's sequence of approximations converges slowly.

6. *Variable relative rate method of solution*

The slow convergence of Picard's sequence is understandable since no matter what y happens to be, the first approximation is assumed to be constant. There is no attempt at fitting this approximation to the form of $f(x,y)$ in (23). Expressing it differently, Picard's first approximation is obtained by taking

$$\frac{dy}{dx} = 0 \tag{33}$$

as the first approximation to (23). We can certainly do better than that.

Let us expand $f(x,y)$ in a power series in the vicinity of $y = b$ by Taylor's formula.

$$f(x,y) = f(x,b) + (y - b)\frac{\partial f}{\partial b} + R(x,y), \tag{34}$$

where $R(x,y)$ is the remainder term or the difference between $f(x,y)$ and the first two terms on the right. Substituting in (23), we have

$$\frac{dy}{dx} = \frac{\partial f}{\partial b} y + \left[f(x,b) - b\frac{\partial f}{\partial b}\right] + R(x,y). \tag{35}$$

If R is neglected, the equation becomes linear.

In order to build a sequence of successive approximations, (35) is treated as an equation of type (1) and the solution is written in the form (20). Since " $Q(\xi)$ " now depends on the unknown function $y(\xi)$ the result does not really represent the solution of the problem. We have merely replaced a differential equation by an integral equation. However, from the new form of the equation we can obtain a rapidly converging series.

Let us see how the process works for equation (30). Since $f(x,y) = y^2$ and $b = 1$, $\partial f/\partial y = 2y$ and $\partial f/\partial b = 2$; also, $f(x,1) = 1$; hence we write (30) as

$$\frac{dy}{dx} = 2y - 1 + (y - 1)^2, \tag{36}$$

where the third term on the right is obtained simply by subtracting the first two from y^2. Using (20) with C equal to unity, we have

$$\begin{aligned} y(x) &= e^{2x} - \int_0^x e^{2(x-\xi)}\,d\xi + \int_0^x [y(\xi) - 1]^2 e^{2(x-\xi)}\,d\xi \\ &= \tfrac{1}{2}(1 + e^{2x}) + e^{2x}\int_0^x [y(\xi) - 1]^2 e^{-2\xi}\,d\xi \end{aligned} \tag{37}$$

The first approximation is obtained by neglecting the last integral

$$\hat{y}_0(x) = \tfrac{1}{2}(1 + e^{2x}) = 1 + x + x^2 + \tfrac{2}{3}x^3 + \cdots. \tag{38}$$

The power series permits us to compare the new method with Picard's. By referring to (31) we find that the present approximation is comparable to (and a little better than) the third of Picard's sequence, $y_2(x)$. Substituting $\hat{y}_0(\xi)$ in the integrand of (37), we obtain

$$\hat{y}_1(x) = \tfrac{3}{8} + \tfrac{1}{2}(1 - x)e^{2x} + \tfrac{1}{8}e^{4x}. \tag{39}$$

The following table presents a comparison between these approximations and the exact solution $y(x) = 1/(1 - x)$.

y \ x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\hat{y}_0(x)$	1.111	1.246	1.411	1.613	1.859	2.160	2.528	2.977	3.525	4.194
$\hat{y}_1(x)$	1.111	1.250	1.428	1.662	1.978	2.417	3.039	3.937	5.279	7.199
$y(x)$	1.111	1.250	1.429	1.667	2.000	2.500	3.333	5.000	10.00	∞

(40)

Normally it is more practical to use only the first approximation over a smaller interval and take the end value of this solution as the initial value for the next interval.

Problems

1. Obtain the linear approximation to (27) and solve it.

Ans. $dy/dx = y\cos(x + 0.5) + [\sin(x + 0.5) - 0.5\cos(x + 0.5)]$,

$$y(x) = 0.5 + e^{\sin(x+0.5)}\int_0^x \sin(\xi + 0.5)\,e^{-\sin(\xi+0.5)}\,d\xi.$$

The integral has to be evaluated numerically.

2. Obtain the solution of

$$\frac{dy}{dx} = A + Bx + ky, \qquad x = 0, \qquad y = 0,$$

where A, B, k are constants. *Ans.* $y = \left(\frac{A}{k} + \frac{B}{k^2}\right)(e^{kx} - 1) - \frac{B}{k}x.$

3. Show that the solution of

$$\frac{dy}{dx} = A + B(x - a) + k(y - b), \qquad x = a, \qquad y = b$$

is $y - b = \left(\frac{A}{k} + \frac{B}{k^2}\right)[e^{k(x-a)} - 1] - \frac{B}{k}(x - a).$

4. Show that in the vicinity of $x = a$, $y = b$ the solution of

$$\frac{dy}{dx} = f(x,y)$$

is given by the formula in Problem 3, where

$$A = f(a,b), \qquad B = \frac{\partial f}{\partial x}\bigg|_{\substack{x=a\\y=b}}, \qquad k = \frac{\partial f}{\partial y}\bigg|_{\substack{y=a\\x=b}}.$$

This " exponential extrapolation formula " may be used for numerical integration of differential equations of the first order.

7. *Linear and nonlinear equations*

There is an important difference between linear and nonlinear equations. The singular points of a linear equation are fixed, that is, they are independent of the initial conditions; furthermore, they are exhibited by the coefficients. If either P or Q is infinite at some point $x = a$, dy/dx is also infinite and $y(x)$ is singular; on the other hand, if P and Q are analytic at $x = a$, we can obtain all derivatives of y by successive differentiation of the equation; the resulting equations indicate clearly that these derivatives exist and that the function is defined in the vicinity of $x = a$. Similarly if either P or Q is many-valued, y is many-valued. Conversely if y is many-valued, either P or Q must be singular.

The singularities of nonlinear equations may be *movable* and they are not necessarily exhibited by the equation itself. Equation (32) represents a particular solution of (30); y becomes infinite at $x = 1$ and there is nothing in (30) to suggest this. The particular solution for which $y = b$ at $x = a$ is

$$x - a = \frac{1}{b} - \frac{1}{y}, \qquad y = \frac{b}{1 + ab - bx}. \tag{41}$$

This solution becomes infinite at a point

$$x = a + \frac{1}{b}, \tag{42}$$

which depends on the initial conditions. This makes the analysis of nonlinear equations vastly more difficult, except in the comparatively few cases in which the solution happens to be composed of a finite number of terms representing known functions. "Comparatively few" refers to equations encountered in applied mathematics and not to manufactured equations set as exercises in texts on differential equations.

Problem. If the initial value of the solution of a first order differential equation is left unspecified, the solution contains an "arbitrary constant." Show that the converse is also true; that is, the family of functions $y = f(x,C)$ depending on an arbitrary parameter C can be described by a first-order differential equation independent of C. First try a special example such as $y = 1/(C - x)$; then present a general argument.

8. *Special methods*

The solution of the most general linear equation of the first order has been reduced to straightforward integration which, however, has to be performed numerically except in certain special cases. Even more often, we have to resort to numerical solution in the case of nonlinear equations. Of course, whether the equation is linear or not, its solution may sometimes be reduced to straightforward integration.

The most obvious cases are equations

$$\frac{dy}{dx} = f(x), \qquad \frac{dy}{dx} = F(y), \tag{43}$$

which, aside from the derivative, contain only one variable. Their solutions are

$$y = b + \int_a^x f(x)\,dx, \qquad x = a + \int_b^y \frac{dy}{F(y)}, \tag{44}$$

or

$$y = \int f(x)\,dx + C, \qquad x = \int \frac{dy}{F(y)} + C, \tag{45}$$

depending on whether the initial condition, $y = b$ when $x = a$, is specified or not.

Step by step, less obvious cases may be similarly treated; thus

$$\frac{dy}{dx} = -\frac{x}{y}, \qquad x\,dx + y\,dy = 0, \qquad x^2 + y^2 = C; \tag{46}$$

or, more generally,

$$\frac{dy}{dx} = \frac{F(y)}{f(x)}, \qquad \frac{dx}{f(x)} = \frac{dy}{F(y)}, \qquad \int \frac{dx}{f(x)} = \int \frac{dy}{F(y)} + C. \tag{47}$$

If $y = b$ when $x = a$,

$$\int_a^x \frac{dt}{f(t)} = \int_b^y \frac{dt}{F(t)}; \tag{48}$$

this form indicates clearly that both variables occur as the limits of integration.

In all the above instances *the variables have been separated.* It may also happen that if the differential equation is expressed in the form

$$M\,dx + N\,dy = 0, \tag{49}$$

where M and N are functions of x and y, the left side is an exact differential of some function $u = f(x,y)$; the solution is then

$$u = f(x,y) = C. \tag{50}$$

For example,

$$y\,dx + x\,dy = 0, \qquad xy = C. \tag{51}$$

In this case the variables can also be separated. If we divide the equation by xy we find: $\log x + \log y = C_1$ or

$$\log(xy) = C_1 \qquad \text{or} \qquad xy = \exp C_1 = C.$$

In Chapter 5 we saw that a differential expression (49) is an exact differential if $\partial N/\partial x = \partial M/\partial y$. When this condition is satisfied, the solution of (49) can be found as follows. We note that

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy, \tag{52}$$

and that the partial differential $M\,dx$ could be derived only from terms containing x; hence,

$$u = \int M\,dx + F(y), \tag{53}$$

where, in the partial integration, y is assumed to be constant. The terms not obtainable by partial integration are those containing y alone. Similarly,

$$u = \int N\,dy + G(x). \tag{54}$$

Now $F(y)$ must be included in the partial integral in (54) and $G(x)$ must

be included in the partial integral in (53); hence we obtain u by taking all terms of one partial integral and adding those of the other integral which do not duplicate the terms already taken. For example,

$$(2 + y)\,dx + (2y + x)\,dy = 0; \quad u = \int (2 + y)\,dx = 2x + yx + F(y),$$

$$u = \int (2y + x)\,dy = y^2 + xy + G(x); \quad u = xy + 2x + y^2 = C. \tag{55}$$

The terms not common to both partial integrals must be functions of a single variable; if we find a term depending on both variables in one partial integral but not in the other, the condition for u being an exact differential is not satisfied. For instance

$$y\,dx - x\,dy = 0, \tag{56}$$

where

$$u = \int y\,dx = xy + F(y), \quad u = -\int x\,dy = -xy + G(x). \tag{57}$$

It is impossible for u to be equal simultaneously to xy and to $-xy$, when the remaining terms depend on only one variable. Since $\partial N/\partial x \neq \partial M/\partial y$, we should not have attempted to treat (56) as an exact differential; still no harm has been done because at a certain stage in the process it becomes clear that the method will not work.

Barring some exceptional points, the differential equation and its derivatives yield all the derivatives of the unknown function and, by Taylor's formula, the solution $y = f(x,C)$, where C is an arbitrary constant. This solution can be extended analytically to all real and complex values of x, again with the exception of certain singular points. Theoretically at least, the equation can be solved for the arbitrary constant $C = g(x,y)$ so that the differential equation of this family of functions would be an exact differential equation

$$\frac{\partial g}{\partial x}\,dx + \frac{\partial g}{\partial y}\,dy = 0. \tag{58}$$

This equation can differ from (49) only by some factor depending on x and y; thus, there must exist an *integrating factor* which would convert a non-exact into an exact differential equation. For instance, dividing (56) by $x^2 + y^2$, we find that the equation becomes exact:

$$\frac{y\,dx - x\,dy}{x^2 + y^2} = d\tan^{-1}(x/y) = 0, \qquad \tan^{-1}(x/y) = C, \tag{59}$$

$$x/y = \tan C = C_1.$$

Similarly, we could divide it by y^2 in which case the left side would become $d(x/y)$; or, we could divide by xy in which case the left side would become $d(\log x - \log y) = d \log (x/y)$.

There are infinitely many integrating factors. To show this we observe that if $g(x,y) = C$, a constant, any function $F[g(x,y)] = F(C) = C_1$ is also a constant; hence

$$dF = \frac{\partial F}{\partial g}\left[\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy\right] = 0 \tag{60}$$

is an exact differential equation for *any* F. However, the problem of finding even one integrating factor is not simple, unless it is obvious by inspection.

It is, however, possible to start with some particular form of the integrating factor and find all equations which can be solved by it. To illustrate, let us find the class of equations for which the integrating factor is a function of x only, $U = U(x)$. Equation (49) is converted into

$$UM\,dx + UN\,dy = 0. \tag{61}$$

If this is to be an exact differential equation, we must have

$$\frac{\partial}{\partial x}(UN) = \frac{\partial}{\partial y}(UM). \tag{62}$$

Remembering that U is a function of x only, we have

$$\frac{dU}{dx}N + U\frac{\partial N}{\partial x} = U\frac{\partial M}{\partial y}, \qquad \frac{1}{U}\frac{dU}{dx} = \frac{1}{N}\left[\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right]. \tag{63}$$

If the right side is a function of x alone, we have a linear differential equation for U and the solution is

$$U = \exp\int\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right)\frac{dx}{N}. \tag{64}$$

In this way numerous rules can be obtained for solving equations of certain forms; these rules can be tabulated and referred to in any special case; but we can hardly expect to make a complete collection of such rules.

Transformations of variables offer unlimited possibilities for the solution of differential equations and exercising one's ingenuity; but there are no straightforward rules which insure success except in cases which have previously been tried out.

Problems

1. Solve the following equations and verify the solutions:

$$(2xy + 7y)\,dx + (x^2 + 7x + 2y)\,dy = 0, \qquad 2x^2y\,dx + dy = 0,$$

$$(x^2 - y^2)\,dx + 2xy\,dy = 0. \qquad \textit{Ans.}\ \ x^2 + y^2 + Cx = 0.$$

2. Under what condition is the integrating factor of (49) a function of $(x+y)$? What is the integrating factor if the condition is satisfied?

Ans. $(D_yM - D_xN)/(N - M) = F(x+y)$, some function of $(x+y)$ only; the integrating factor is $\exp \int F(x+y)d(x+y)$.

3. Solve (27) by introducing a new variable $u = x + y$.

Ans. $x + 2/[1 + \tan \frac{1}{2}(x+y)] = C$, or $y = -x + 2 \tan^{-1} [(2 - C + x)/(C - x)]$.

4. Start with some function of x, y, C; differentiate and eliminate C; then try to "solve" the differential equation.

5. Consider a pendulum of length l and assume that the mass of the rod is negligible compared with the mass M of the bob. If $\theta(t)$ is the angle between the downward vertical and the direction of the rod at the instant $t = t$, and if the energy dissipation is neglected, the equation of motion is

$$Mgl(1 - \cos \theta) + \frac{1}{2} Ml^2 \left(\frac{d\theta}{dt}\right)^2 = E,$$

where E is the total energy of the pendulum and g is the gravitational constant.

Find the maximum angle of deflection; find an approximate solution of the equation and the period of oscillation when θ is small; find the exact solution.

CHAPTER XI

Differential Equations of the Second Order

It would be difficult to overemphasize the practical importance of differential equations of the second order, when we consider the frequency with which they appear in various problems. To treat the subject fully would require a book of substantial size, and in this chapter we restrict ourselves to linear equations; even so, space considerations preclude anything approaching a complete account. We can give only an introduction to the theory and some of its more immediately useful applications.

1. *Homogeneous equations with constant coefficients*

The simplest equation of the second order is

$$\frac{d^2y}{dx^2} + p\frac{dy}{dx} + qy = 0, \tag{1}$$

where p and q are constants. There is no loss of generality in the assumption that the coefficient of the second-order derivative is unity. If the equation is to be of the second order this coefficient must differ from zero, and the equation may be divided by it.

Knowing that the derivatives of exponential functions are proportional to the functions themselves, we assume a solution of the form

$$y = Ae^{\gamma x}, \tag{2}$$

and substitute it in (1):

$$A\gamma^2e^{\gamma x} + Ap\gamma e^{\gamma x} + Aqe^{\gamma x} = 0.$$

The common factor $A \exp(\gamma x)$ can be canceled and the following *characteristic equation* obtained

$$\gamma^2 + p\gamma + q = 0. \tag{3}$$

Thus A may be completely arbitrary; but γ must satisfy the characteristic equation and assume one of the "characteristic" or "proper" values. These values are

$$\gamma_1 = -\tfrac{1}{2}p + \sqrt{\tfrac{1}{4}p^2 - q}, \qquad \gamma_2 = -\tfrac{1}{2}p - \sqrt{\tfrac{1}{4}p^2 - q}. \tag{4}$$

The sum $y_1 + y_2$ of two solutions is also a solution; this may be verified

by substitution. In fact, any linear function $ay_1 + by_2$ is a solution. Hence

$$y = A_1e^{\gamma_1 x} + A_2e^{\gamma_2 x} \tag{5}$$

is a solution of (1).

The most general solution of (1) should contain two arbitrary constants. The equation gives merely a relationship between the function and its first two derivatives. At a given point any two of them can be assigned almost arbitrary values; exceptions arise only when either p or q is equal to zero. Since the coefficient of the second-order derivative is different from zero, y and y' may be assigned arbitrarily at the point in question without any exceptions; y'' is then determined uniquely. By successive differentiation of (1) we may express all higher derivatives in terms of y and y' at a given point. Applying Taylor's theorem we shall get the most general *analytic* solution of the differential equation.

We shall be able to claim that (5) is the most general analytic solution if we can show that at some point, $x = 0$ for instance, we can express A_1 and A_2 in terms of arbitrarily assigned values of $y(0)$ and $y'(0)$. We leave the proof to the reader. The only exception is an obvious one: if $\gamma_1 = \gamma_2$, we have in effect only one arbitrary constant, $A_1 + A_2 = y(0)$. If $y(0)$ is assigned arbitrarily, only one constant in the set (A_1,A_2) is arbitrary.

To find out what happens in the exceptional case $\gamma_1 = \gamma_2$, let us examine the case in which γ_1 and γ_2 are *nearly* equal. After all, we have *two* independent solutions so long as the characteristic values are different, even though the difference may be very small. Now the difference $\exp \gamma_1 x - \exp \gamma_2 x$ is a solution; so is the ratio

$$y = \frac{\exp(\gamma_1 x) - \exp(\gamma_2 x)}{\gamma_1 - \gamma_2}. \tag{6}$$

As γ_1 and γ_2 approach the single characteristic value γ, we expect the limit

$$y = \frac{\partial}{\partial\gamma}\exp(\gamma x) = xe^{\gamma x} \tag{7}$$

to be a solution. Let us make sure of it.

The case $\gamma_1 = \gamma_2 = -p/2$ arises when $q = \frac{1}{4}p^2$ and (1) is

$$y'' + py' + \tfrac{1}{4}p^2y = 0. \tag{8}$$

Substituting from (7), we find that we really have found another solution in addition to $\exp(\gamma x)$. The general solution becomes

$$y = A_1e^{\gamma x} + A_2xe^{\gamma x}, \qquad \gamma = -p/2. \tag{9}$$

For some applications of (1) and for a more detailed discussion of its solutions the reader is referred to texts on mechanical and electrical oscillations.

Problems

1. Solve the following equations

$$\begin{array}{ll} \text{a) } y'' - 5y' + 6y = 0, & \text{b) } y'' + 5y' + 6y = 0, \\ \text{c) } y'' + y' - 6y = 0, & \text{d) } y'' + 6y' + 9y = 0, \\ \text{e) } y'' + y = 0, & \text{f) } y'' + 2y' + 5y = 0. \end{array}$$

2. In the case (e) of the preceding problem the exponential solution is complex; reduce it to the real form $y = A \cos x + B \sin x = M \cos (x + \varphi)$ by an appropriate change in the constants of integration, noting that these constants do not have to be real.

3. Show that in the case (f) of Problem 1, the solution can be expressed as $y = e^{-x} (A \cos 2x + B \sin 2x) = Me^{-x} \cos (2x + \varphi)$.

4. Show that if $q > \frac{1}{4}p^2$, the solutions of (1) are oscillatory. The amplitude of oscillation either decreases or increases with x according as p is positive or negative. Show that if p is small and q is positive, the frequency of oscillations is determined largely by q. Show also that if $q < 0$, the solutions are nonoscillatory (assuming that p is real).

5. Find solutions of the equations in Problem 1 subject to the following "initial" or "boundary" conditions:

$$\begin{array}{lll} \text{A) } y(0) = 1, & y'(0) = 2; & \\ \text{B) } y(1) = 1, & y'(1) = 2; & -3\ 6 \\ \text{C) } y(0) = 1, & y(1) = 2; & \\ \text{D) } y(0) = 1, & y'(1) = 1. & \end{array}$$

6. Prove directly from (5) and (9) that $y(a)$ and $y'(a)$ can be assigned arbitrarily. *Hint:* Recall the conditions under which a system of two linear algebraic equations can or cannot be solved.

7. Prove that $y(a)$ and $y(b)$ can be assigned almost arbitrarily. Show that if $\gamma_1 \neq \gamma_2$ and $(\gamma_1 - \gamma_2)(a - b) = 2n\pi i$, where n is an integer, then $y(a)$ and $y(b)$ cannot be assigned arbitrarily. *Hint:* Proceed boldly with the determination of the arbitrary constants and note that the results become meaningless if the denominators vanish, unless the numerators also vanish.

8. Show that if the conditions of the preceding problem are satisfied, there are some pairs $y(a)$, $y(b)$ for which infinitely many solutions exist (because one constant of integration could be assigned an arbitrary value). Find these pairs $y(a)$, $y(b)$. *Ans.* $y(b) = y(a) \exp [\gamma_1(b - a)] = y(a) \exp [\gamma_2(b - a)]$.

9. Solve (1) subject to $y(0) = a$, $y'(0) = b$ by Maclaurin's formula.

10. Assume a solution of (1) in the form of a power series and obtain equations for the coefficients. Note that two coefficients are not determined by the equation.

2. *Basic sets of solutions*

Since a linear combination of two particular solutions, $y_1(x)$ and $y_2(x)$, of (1) is also a solution, the general solution may be conveniently expressed in the following form

$$y(x) = y(a)y_1(x) + y'(a)y_2(x), \tag{10}$$

where

$$\begin{aligned} y_1(a) &= 1, \qquad y_1'(a) = 0; \\ y_2(a) &= 0, \qquad y_2'(a) = 1. \end{aligned} \tag{11}$$

Such a pair of solutions is said to constitute a *basic set.*

Problems

1. Show that if the roots of the characteristic equation are distinct, a basic set with $y_1(0) = 1$, $y_2'(0) = 1$, is

$$y_1(x) = (\gamma_2 e^{\gamma_1 x} - \gamma_1 e^{\gamma_2 x})/(\gamma_2 - \gamma_1), \qquad y_2(x) = (e^{\gamma_1 x} - e^{\gamma_2 x})/(\gamma_1 - \gamma_2). \tag{12}$$

2. If the roots of the characteristic equation are equal, the corresponding basic set is

$$y_1(x) = e^{\gamma x}, \qquad y_2(x) = xe^{\gamma x}. \tag{13}$$

3. What are the fundamental sets with reference to $x = a$?

3. *Nonhomogeneous equations with constant coefficients*

The next equation to be considered is

$$y'' + py' + qy = f(x), \tag{14}$$

where $f(x)$ is a given function. If we can find some function $\hat{y}(x)$ which is a particular solution of this equation and add to it the general solution $\bar{y}(x)$ of the corresponding homogeneous equation, in which $f(x) = 0$, we obtain the general solution of the nonhomogeneous equation,

$$y(x) = \bar{y}(x) + \hat{y}(x). \tag{15}$$

To prove this we substitute in (14) and use the assumed properties of $\hat{y}$ and $\bar{y}$. In this connection $\bar{y}(x)$ is called the *complementary function.*

At times the form of $f(x)$ suggests a possible particular solution. If $f(x) = A \exp(kx)$, our knowledge of exponential functions suggests that $\hat{y}(x)$ might be of the form $B \exp(kx)$; substituting in the equation, we find

$$\begin{aligned} Bk^2e^{kx} + Bpke^{kx} + Bqe^{kx} &= Ae^{kx}, \\ B &= A/(k^2 + pk + q). \end{aligned} \tag{16}$$

The assumption has been verified and the unknown constant B determined, provided k does not equal either of the two characteristic values. When k is equal to one of these values, the denominator becomes infinite and the assumed form does not yield a solution. In this case, we try $\bar{y}(x) = Bx \exp(kx)$; substituting in the equation we find that $\bar{y}(x)$ is a solution if $B = A/(2k + p)$.

When we are unable to obtain a particular solution by inspection, we can turn to a general method which is analogous to the one developed in the preceding chapter for equations of the first order. Let $f(x)$ equal zero except in a vanishingly small neighborhood of $x = \xi$. We shall denote this interval by $(\xi - 0, \xi + 0)$ with the understanding that the interval is $(\xi - \delta, \xi + \delta)$ where δ approaches zero. Integrating (14) in this interval we have

$$[y'(\xi + 0) - y'(\xi - 0)] + p[y(\xi + 0) - y(\xi - 0)] + q\int_{\xi-0}^{\xi+0} y(x)\,dx = \int_{\xi-0}^{\xi+0} f(x)\,dx. \tag{17}$$

We assume that $f(x)$ increases as δ approaches zero in such a way that the integral on the right remains unity

$$\int_{\xi-0}^{\xi+0} f(x)\,dx = 1. \tag{18}$$

Next we assume that $y(x)$ is finite and continuous at $x = \xi$, in which case (17) becomes

$$y'(\xi + 0) - y'(\xi - 0) = 1. \tag{19}$$

Under the stated conditions $y(x)$ is such that its derivative is discontinuous at $x = \xi$ and the increment is unity. Thus we have

$$\begin{aligned} y(x) &= y(a)y_1(x) + y'(a)y_2(x), & x < \xi, \\ y(x) &= y(a)y_1(x) + y'(a)y_2(x) + G(x,\xi), & x > \xi, \end{aligned} \tag{20}$$

as the general solution of (14) for the function $f(x)$ which vanishes outside a vanishingly small neighborhood of $x = \xi$, while within the neighborhood it is infinite in such a way that the integral (18), the *moment* of $f(x)$ in the interval $(\xi - 0, \xi + 0)$, is unity. In equation (20) y_1 and y_2 constitute a basic set (11) of solutions of the associated homogeneous equation; while *Green's function* $G(x,\xi)$ is that solution of the homogeneous equation which vanishes at $x = \xi$ and whose derivative there is unity; that is

$$G(\xi,\xi) = 0, \qquad G'(\xi,\xi) = 1, \tag{21}$$

where the prime indicates the derivative of G with respect to x.

To construct the solution of (14) for any $f(x)$, we apply the superposition principle, that is, we regard $f(x)$ as a sum of functions of the impulse type with moments equal to $f(\xi)\,d\xi$; thus

$$\begin{aligned} y(x) &= y_0(x) + \int_a^x G(x,\xi)f(\xi)\,d\xi, \\ y_0(x) &= y(a)y_1(x) + y'(a)y_2(x). \end{aligned} \tag{22}$$

To verify that this is really the general solution of (14), we differentiate twice with respect to x (remembering that x occurs in the integrand and in the upper limit) and use (21):

$$\begin{aligned} y'(x) &= y_0'(x) + \int_a^x G'(x,\xi)f(\xi)\,d\xi + G(x,x)f(x) \\ &= y_0'(x) + \int_a^x G'(x,\xi)f(\xi)\,d\xi, \\ y''(x) &= y_0''(x) + \int_a^x G''(x,\xi)f(\xi)\,d\xi + G'(x,x)f(x) \\ &= y_0''(x) + \int_a^x G''(x,\xi)f(\xi)\,d\xi + f(x). \end{aligned} \tag{23}$$

Thus the left side of (14) becomes

$$\begin{aligned} y'' + py' + qy = y_0'' &+ py_0' + qy_0 + f(x) \\ &+ \int_a^x [G''(x,\xi) + pG'(x,\xi) + qG(x,\xi)]f(\xi)\,d\xi. \end{aligned} \tag{24}$$

The first three terms on the right reduce to zero because $y_0(x)$ is a solution of the homogeneous equation; for the same reason the bracketed factor in the integrand vanishes; only $f(x)$ remains and the proof has been completed.

The particular solution represented by the integral in (22) vanishes at $x = a$; as seen from (23), its derivative also vanishes at $x = a$; thus the initial conditions at $x = a$ enter entirely through the complementary function $y_0(x)$.

For example, let us solve

$$y'' + \beta^2 y = f(x). \tag{25}$$

The associated homogeneous equation is

$$y_0'' + \beta^2 y_0 = 0. \tag{26}$$

If $y(a)$ and $y'(a)$ are the initial values, then

$$y_0(x) = y(a)\cos\beta(x-a) + \frac{1}{\beta}y'(a)\sin\beta(x-a). \tag{27}$$

That $\cos\beta(x-a)$ and $\sin\beta(x-a)$ are solutions of (26) may be found either from the general results of Section 1 or directly by inspection. Green's function is

$$G(x,\xi) = \frac{1}{\beta}\sin\beta(x-\xi); \tag{28}$$

it satisfies (26), it vanishes at $x = \xi$ and its derivative there is unity. The complete solution is

$$\begin{aligned} y(x) &= y_0(x) + \frac{1}{\beta}\int_a^x \sin\beta\,(x-\xi)f(\xi)\,d\xi \\ [&= y_0(x) + \frac{1}{\beta}\sin\beta x\int_a^x f(\xi)\cos\beta\xi\,d\xi \\ &\quad - \frac{1}{\beta}\cos\beta x\int_a^x f(\xi)\sin\beta\xi\,d\xi. \end{aligned} \tag{29}$$

In this form of the solution $f(x)$ is not required to be analytic; for example, $f(x)$ may equal $\sin x$ in the interval $(0,\pi)$ and be zero everywhere else. Likewise, $f(x)$ might represent experimental data in which case the integrals would have to be evaluated numerically.

Problems

1. Find by inspection particular solutions of

$y'' + 4y = 12, \qquad y'' + y = 2x, \qquad y'' + y' + y = x.$

Ans. $3, \quad 2x, \quad x - 1.$

2. Solve $y'' + y = f(x)$, where $f(x)$ is unity in the interval $(0,x_0)$ and zero elsewhere; $y(0) = 1$ and $y'(0) = 0$.

Ans.
$$\begin{aligned} y(x) &= \cos x, & x &< 0; \\ &= 1, & 0 &< x < x_0; \\ &= \cos(x - x_0), & x &> x_0. \end{aligned}$$

3. Solve the preceding problem when $f(x) = 2$ in the interval $(0,x_0)$ and zero elsewhere.

Ans.
$$\begin{aligned} y(x) &= \cos x, & x &< 0; \\ &= 2 - \cos x, & 0 &< x < x_0; \\ &= 2\cos(x - x_0) - \cos x, & x &> x_0. \end{aligned}$$

4. Solve Problem 2 when $f(x) = \cos x$ in the interval $(0,x_0)$ and zero elsewhere.

Ans. $y(x) = \cos x, \qquad x < 0;$

$$= \cos x + \tfrac{1}{2}x \sin x, \qquad 0 < x < x_0;$$

$$= \tfrac{3}{4}\cos x + \tfrac{1}{2}x_0 \sin x + \tfrac{1}{4}\cos(x - 2x_0), \qquad x > x_0.$$

5. *What are the expressions for $y_0(x)$ and $G(x,\xi)$ in the general case* (14)?

Ans.
$$y_0(x) = y(a)\,\frac{\gamma_2 \exp\gamma_1(x-a) - \gamma_1 \exp\gamma_2(x-a)}{\gamma_2 - \gamma_1}$$
$$+ y'(a)\,\frac{\exp\gamma_1(x-a) - \exp\gamma_2(x-a)}{\gamma_1 - \gamma_2};$$
$$G(x,\xi) = \frac{\exp\gamma_1(x-\xi) - \exp\gamma_2(x-\xi)}{\gamma_1 - \gamma_2}.$$

Note that $G(x,\xi)$ is obtained from $y_0(x)$ if we let $y(a) = 0$, $y'(a) = 1$, and replace a by ξ.

4. *A physical interpretation*

The " shock method " of solution of nonhomogeneous equations of the second order may be illustrated by a physical example (cf. Section 3 of the last chapter). Consider the displacement y of a mass M attached to a spring with stiffness S under the influence of a force $F(t)$, Figure 1.16. Assuming that motion takes place in a viscous medium offering a resistance proportional to the velocity, and that displacements are small enough for elastic forces to be proportional to them, we obtain the following equation of motion

$$M\ddot{y} + R\dot{y} + Sy = F(t). \tag{30}$$

Each dot superscript in this equation represents differentiation with respect to time t.

In this case the integral (18) represents the impulse of force. A continuously acting force may be regarded as the limit of a succession of small impulses $F(t)\Delta t$; thus the response to this force may be obtained by integrating the response to a typical shock if we assume that the responses to the several shocks are mutually independent. The assumption is justified if the equation, and hence the mechanical system, is linear. Thus from three given displacements, $y_1(t)$, $y_2(t)$ and $y_1(t) + y_2(t)$, we obtain from (30) the required forces $F_1(t)$, $F_2(t)$ and $F_1(t) + F_2(t)$. Thus we automatically obtain a response $y_1 + y_2$ to the force $F_1 + F_2$. Experience shows that the response to any force is unique and thus we can be sure of obtaining $y_1 + y_2$ from $F_1 + F_2$ as well as $F_1 + F_2$ from $y_1 + y_2$. To restate the argument: if $y(t)$ is another response to $F_1 + F_2$, then by sub-

tracting the corresponding equation, we find that $y - (y_1 + y_2)$ is a response to a force equal to zero *at all times.* It is safe to assume that this response is also zero.

The response to a unit impulse of force at $t = \xi$ will depend only on the interval $t - \xi$ which has elapsed after the impulse has been applied, and we may denote it by $K(t - \xi)$; this function vanishes if $t \leq \xi$. In view of the preceding argument we have

$$y(t) = \int_0^t K(t - \xi)F(\xi)\, d\xi. \tag{31}$$

Green's function $G(t,\xi)$ of the preceding section now appears as $K(t - \xi)$.

To find Green's function we note that the impulse of force equals the increase in the momentum $M\dot{y}$. Hence the unit impulse acting on a body at rest communicates to it a velocity $1/M$. After the shock we have $F(t) = 0$ in (30), and $K(t - \xi)$ is that solution of the homogeneous equation which vanishes at $t = \xi$ and whose derivative at that moment is $1/M$.

In an electric circuit consisting of an inductance "M," resistance "R" and capacitance "$1/S$," *all in series*, y is the charge passing through the circuit, $\dot{y}$ is the electric current, and F is the impressed voltage. A voltage impulse is equal to a sudden increase in the magnetic flux through the inductor. In an electric circuit consisting of a capacitance "M," conductance "R" and inductance "$1/S$," *all in parallel*, y is the voltage across the circuit and hence across the capacitor; F is the current entering the circuit. In this case the impulse of F equals the charge suddenly transferred from one plate of the capacitor to the other.

Problem. What is the response of the mechanical system considered in this section to a unit impulse of force administered at $t = 0$?

Ans. $$K(t) = \frac{\exp(-Rt/2M)\sin t\sqrt{(S/M) - (R^2/4M^2)}}{M\sqrt{(S/M) - (R^2/4M^2)}}$$

provided $2\sqrt{SM} > R$ and the square root is real; otherwise

$$K(t) = \frac{e^{p_1 t} - e^{p_2 t}}{M(p_2 - p_1)}, \qquad p_{1,2} = -\frac{R}{2M} \pm \sqrt{\frac{R^2}{4M^2} - \frac{S}{M}}.$$

5. *Linear equations with variable coefficients*

The most general linear equation with variable coefficients is

$$\frac{d^2y}{dx^2} + p(x)\frac{dy}{dx} + q(x)y = f(x). \tag{32}$$

The equation is homogeneous if $f(x) = 0$. Some equations of this type play such important roles in applied mathematics that they have acquired

distinguishing names; for instance,

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+(x^2-n^2)y=0 \quad \text{(Bessel's equation);}$$

$$\sin\theta\frac{d^2y}{d\theta^2}+\cos\theta\frac{dy}{d\theta}+n(n+1)\sin\theta y=0 \quad \text{(Legendre's equation).} \tag{33}$$

Just as in the case of equations with constant coefficients, the solution of the general equation can be expressed in the form (22) where $y_1(x)$ and $y_2(x)$ form a basic set (11) of solutions of the associated homogeneous equation. Green's function $G(x,\xi)$ is also a solution of the homogeneous equation, subject to conditions (21). The arguments leading to this solution depend solely on the linearity of the equation with respect to y and its derivatives; hence the solution of the homogeneous case constitutes the main problem.

There is only one limitation on the generality of the preceding statement: $p(x)$ and $q(x)$ must obviously exist at $x = a$ or we should be unable to assign arbitrary values to the function and its derivative without causing the second derivative to become infinite. Such exceptional points are called the *singular points* of the differential equation; all points at which p and q are analytic are *ordinary points*.

Even if $x = a$ is a singular point the solution of the nonhomogeneous equation may often be expressed in the form (22), with $y_0(x)$ equal to any solution of the associated homogeneous equation. The condition is that all the operations needed in the verification of (22) should be permissible. It is easier to establish whether or not this condition is fulfilled in each special case rather than to try to obtain general rules and remember them.

6. *Expansions in power series*

It would be easy to construct hundreds of linear equations whose solutions could be expressed as combinations of a finite number of "elementary" functions, that is, algebraic, circular, exponential and hyperbolic functions. For instance, starting with $y = A_1f_1(x) + A_2f_2(x)$ where f_1 and f_2 are known, such equations would be obtained by eliminating the arbitrary constants from the equations for y and its first two derivatives. The solutions of some very complicated looking equations are expressible in terms of known functions; but as a general rule these equations will be found only in sets of exercises for students and do not occur in practical applications. Even such simple equations as (33) cannot be solved in terms of elementary functions, and these equations should be regarded as definitions of the functions which satisfy them.

It is usually possible, however, to express any solution in terms of an

infinite number of elementary functions, powers of x for instance. Consider the Bessel equation and assume

$$y = x^m(a_0 + a_1x + a_2x^2 + \cdots). \tag{34}$$

Differentiating, we have

$$\begin{aligned}
y' &= a_0mx^{m-1} + a_1(m+1)x^m + a_2(m+2)x^{m+1} + \cdots, \\
y'' &= a_0m(m-1)x^{m-2} + a_1(m+1)mx^{m-1} \\
&\qquad + a_2(m+2)(m+1)x^m + \cdots, \\
x^2y'' &= a_0m(m-1)x^m + a_1(m+1)mx^{m+1} \\
&\qquad + a_2(m+2)(m+1)x^{m+2} + \cdots, \\
xy' &= a_0mx^m + a_1(m+1)x^{m+1} + a_2(m+2)x^{m+2} + \cdots, \\
x^2y &= a_0x^{m+2} + a_1x^{m+3} + \cdots, \\
-n^2y &= -a_0n^2x^m - a_1n^2x^{m+1} - a_2n^2x^{m+2} - a_3n^2x^{m+3} - \cdots.
\end{aligned} \tag{35}$$

Substituting in the Bessel equation and equating to zero the coefficients of successive powers of x, we obtain

$$a_0(m^2 - n^2) = 0, \qquad a_1[(m+1)^2 - n^2] = 0, \tag{36}$$

and

$$\begin{aligned}
a_2[(m+2)^2 - n^2] + a_0 &= 0, \\
a_3[(m+3)^2 - n^2] + a_1 &= 0, \\
\cdots\cdots\cdots\cdots & \\
a_k[(m+k)^2 - n^2] + a_{k-2} &= 0.
\end{aligned} \tag{37}$$

The first coefficient a_0 must be different from zero if an expansion of the form (34) is to be possible; for a_0 is the first nonvanishing coefficient whatever it may be. Hence, from (36),

$$m^2 - n^2 = 0, \qquad m = \pm n. \tag{38}$$

For either of these values of m, $a_1 = 0$ since $(\pm n + 1)^2 - n^2 \neq 0$; hence all a_k vanish for odd values of k. For the even values of k we obtain successively

$$a_2 = -\frac{a_0}{[(\pm n + 2)^2 - n^2]}, \qquad a_4 = -\frac{a_2}{[(\pm n + 4)^2 - n^2]},$$

$$a_{2p} = -\frac{a_{2p-2}}{[(\pm n + 2p)^2 - n^2]}. \tag{39}$$

Taking the upper sign, we find that $a_{2p} = a_{2p-2}/4p(n+p)$ and the

solution is

$$y_1(x) = a_0\left(x^n - \frac{x^{n+2}}{4(n+1)} + \frac{x^{n+4}}{4^2 1\cdot 2(n+1)(n+2)} - \cdots\right)$$

$$= a_0 n! \sum_{m=0}^{\infty} \frac{(-)^m x^{n+2m}}{m!\,(n+m)!\,2^{2m}}. \tag{40}$$

The second solution is obtained from (40) by changing the sign of n. There is no difficulty about it in the expanded form of the series; in the condensed form, it is necessary to define the meaning of factorials of negative numbers (see the Chapter on the gamma function).

Since the general solution of a homogeneous equation is a linear combination of two independent particular solutions, we are free to choose the coefficient a_0 to suit our purposes. The standard set of *Bessel functions* of order n is

$$\begin{aligned} J_n(x) &= \sum_{m=0}^{\infty} \frac{(-)^m x^{n+2m}}{m!\,(n+m)!\,2^{n+2m}}, \\ J_{-n}(x) &= \sum_{m=0}^{\infty} \frac{(-)^m x^{-n+2m}}{m!\,(-n+m)!\,2^{-n+2m}}. \end{aligned} \tag{41}$$

On a closer inspection of the expanded form in (40) we find that we have two solutions *only if n is not an integer*. If n is a negative integer, sooner or later we reach a term with a zero denominator and have to rule out the series. Thus if n is an integer, only one solution of Bessel's equation can be expressed as a power series of the form (34). In this case the two Bessel functions defined by (41) become identical, except for the factor $(-)^n$, because the factorials of the negative integers are infinite. The series in powers of $(x - a)$ automatically restricts the behavior of the function it represents, and rules out some functions such as $\log (x - a)$.

The second solution for integral values of n can be found by the method explained in Section 1 in connection with equal roots of the characteristic equation. The two functions $J_{n+\delta}(x)$ and $(-)^n J_{-n-\delta}(x)$ approach equality as δ approaches zero; if their difference divided by 2δ approaches a limit, the limit is likely to be a second solution. We could calculate the limit and verify our anticipation by substitution in the equation.

A similar method consists in introducing the following solution of Bessel's equation for nonintegral values of the parameter

$$N_n(x) = \frac{J_n(x)\cos n\pi - J_{-n}(x)}{\sin n\pi}. \tag{42}$$

It is just a linear combination of two already established solutions so

constructed that it continues to exist and approaches a limit as n approaches an integral value.

Both methods depend essentially on obtaining the derivative $\partial J_n/\partial n$ of a Bessel function with respect to the parameter n. The custom of writing n as a subscript should not blind us to the fact that n is really another variable; the symbol $J(x,n)$ would have exhibited this feature in a conventional way. In obtaining the derivative we have to differentiate x^n with respect to n; since $x^n = \exp(n \log x)$, the derivative is $(\log x) \times \exp(n \log x) = x^n \log x$; the logarithmic factor explains the failure of (34) to represent this particular solution. For further details the reader is referred to the Chapter on Bessel functions.

The principal weakness of power series is their slow convergence except in a limited range of the variable, even when the series converge everywhere. It is usually necessary to supplement power series with other representations.

Problems

1. Show that $x^2y'' + pxy' + qy = 0$ where p and q are constants, usually possesses solutions of the form $y = Ax^n$ and find the proper values of n. What is the second solution when the proper values are equal?

Ans. $n = \frac{1}{2}(1-p) \pm [\frac{1}{4}(1-p)^2 - q]^{1/2}$; if $q = \frac{1}{4}(1-p)^2$, then the solutions are $A_1x^n + A_2x^n \log x$, where $n = \frac{1}{2}(1-p)$.

2. What happens if n is a pure imaginary or complex?

Ans. The general solution may be expressed as $x^{n_1}[A \cos(n_2 \log x) + B \sin(n_2 \log x)]$ where $n = n_1 + in_2$ is given as before.

3. Obtain power series solutions of $y'' = -xy$.

4. Introduce a new independent variable, defined by $x = \cos\theta$, into Legendre's equation (33).

Ans. $(1-x^2)y'' - 2xy' + n(n+1)y = 0$.

5. Obtain power series for the solutions of Legendre's equation and show that if n is an integer one of the series is a polynomial.

7. *Elimination of the term containing the first derivative*

The following substitution

$$y(x) = u(x)v(x), \tag{43}$$

where

$$u(x) = \exp\left[-\tfrac{1}{2}\int^x p(x)\,dx\right], \tag{44}$$

reduces (32) to an equation free from the first derivative

$$v'' + (q - \tfrac{1}{4}p^2 - \tfrac{1}{2}p')v = f(x)\exp\left(\tfrac{1}{2}\int^x p\,dx\right). \tag{45}$$

To prove this theorem we substitute (43) in (32) and collect terms as follows

$$uv'' + (2u' + pu)v' + (u'' + pu' + qu)v = f(x). \tag{46}$$

We are free to choose one of the factors in (43); we choose u to make the coefficient of v' zero;

$$2u' + pu = 0, \qquad u' = -\tfrac{1}{2}pu. \tag{47}$$

This is a linear homogeneous equation of the first order and its solution is (44). The lower limit can be assigned arbitrarily but the same limit must be used in (44) and (45). To complete the reduction, (44) is differentiated twice; then it is found that the exponential factor disappears from the ratio $(u'' + pu' + qu)/u$.

Problems

1. Show that solutions of Bessel's equation are $y = x^{-1/2}v$, where v is a solution of $v'' = -\left(1 - \dfrac{n^2 - 0.25}{x^2}\right)v$. For sufficiently large x, $v \simeq A\cos x + B\sin x$.

2. Show that solutions of Legendre's equation (33) are $y = v/\sqrt{\sin\theta}$, where $v'' = -[(n + \frac{1}{2})^2 + \frac{1}{4}\csc^2\theta]v$. For sufficiently large n, $v \simeq A\cos(n + \frac{1}{2})\theta + B\sin(n + \frac{1}{2})\theta$ provided $\sin\theta$ is not too small.

8. *Normalization of the coefficient of the dependent variable*

Introducing a new independent variable

$$\varphi = \int_a^x \sqrt{q(x)}\,dx, \qquad d\varphi/dx = \sqrt{q(x)}, \tag{48}$$

into (32), we obtain an equation in which the coefficient of y is unity

$$\frac{d^2y}{d\varphi^2} + \left(\frac{p}{\sqrt{q}} + \frac{1}{2q}\frac{dq}{d\varphi}\right)\frac{dy}{d\varphi} + y = \frac{f[x(\varphi)]}{q}. \tag{49}$$

In the new equation p, q, f are expressed as functions of φ.

Since

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{d\varphi}\frac{d\varphi}{dx} = \frac{dy}{d\varphi}\sqrt{q},\\ \frac{d^2y}{dx^2} &= \frac{d^2y}{d\varphi^2}q + \frac{dy}{d\varphi}\frac{d}{dx}\sqrt{q} = \frac{d^2y}{d\varphi^2}q + \frac{dy}{d\varphi}\left(\frac{d}{d\varphi}\sqrt{q}\right)\frac{d\varphi}{dx}\\ &= q\frac{d^2y}{d\varphi^2} + \frac{1}{2}\frac{dq}{d\varphi}\frac{dy}{d\varphi},\end{aligned} \tag{50}$$

equation (32) becomes (49).

If $p = -(dq/d\varphi)/2\sqrt{q} +$ constant, the coefficients in (49) are constants and the general solution may be expressed in terms of integrals.

This transformation is suggested by the following considerations. If β is constant, solutions of $y'' + \beta^2 y = 0$ are $\exp(\pm i\beta x)$; in a small interval the solutions will be approximately exponential even if β is variable; proceeding by infinitesimal steps, we expect approximate solutions in the form $\exp\left(\pm i\int \beta\, dx\right)$; to find out what really happens we introduce a new variable as in (48). If the reader is familiar with the physical behavior of electrical (or mechanical) transmission lines, he will find, as explained in Section 12, that this transformation and the one in the preceding section possess simple interpretations.

Problems

1. Reduce the coefficient of y in $y'' + xy = 0$ to unity.

Ans. $\varphi = \frac{2}{3}x\sqrt{x}, \qquad y''(\varphi) + \left(\frac{1}{3\varphi}\right)y'(\varphi) + y(\varphi) = 0.$

2. In the answer to the preceding problem reduce the coefficient of the first derivative to zero.

Ans. $y(\varphi) = \varphi^{-1/6}v(\varphi), \qquad \text{where } v''(\varphi) + \left(1 + \dfrac{5}{36\varphi^2}\right)v(\varphi) = 0.$

3. Combining the results of the preceding problems, show that for large x the approximate solution of $y'' + xy = 0$ is $y(x) \simeq x^{-1/4}[A\cos(\frac{2}{3}x\sqrt{x}) + B\sin(\frac{2}{3}x\sqrt{x})]$. How large should x be for this approximation to be reasonably accurate?

4. Reduce the coefficients of y in $y'' + x^n y = 0$ and $y'' - x^n y = 0$ to unity in the first equation and negative unity in the second.

Ans. $\varphi = [2/(n+2)]x^{(n+2)/2}, \qquad y''(\varphi) + \dfrac{n}{(n+2)\varphi}y'(\varphi) + y(\varphi) = 0,$

$u = [2/(n+2)]x^{(n+2)/2}, \qquad y''(u) + \dfrac{n}{(n+2)u}y'(u) - y(u) = 0.$

5. Reduce the coefficients of the first derivatives in the equations of the preceding problem to zero.

Ans. $y(\varphi) = \varphi^{-n/2(n+2)}v(\varphi), \qquad v''(\varphi) + \left[1 + \dfrac{n(n+4)}{4(n+2)^2\varphi^2}\right]v(\varphi) = 0,$

$y(u) = u^{-n/2(n+2)}v(u), \qquad v''(u) + \left[-1 + \dfrac{n(n+4)}{4(n+2)^2u^2}\right]v(u) = 0.$

6. From the results of Problems 4 and 5 show that for large x the approximate

solutions of the equations $y'' + x^n y = 0$ and $y'' - x^n y = 0$ are respectively

$$y(x) \simeq x^{-n/4}\left[A\cos\frac{2x^{(n+2)/2}}{n+2} + B\sin\frac{2x^{(n+2)/2}}{n+2}\right],$$

$$y(x) \simeq x^{-n/4}\left[A\exp\frac{2x^{(n+2)/2}}{n+2} + B\exp\left(-\frac{2x^{(n+2)/2}}{n+2}\right)\right].$$

9. *The nature of the solutions of linear homogeneous equations*

The function $f(x)$ in the nonhomogeneous equation affects its solutions to such an extent that nothing can be said about their behavior without specifying $f(x)$. In the homogeneous case the situation is different; by inspecting p and q, it is possible to state whether or not the solution tends to be oscillatory or nonoscillatory. If the range of x is subdivided into small intervals, then in most intervals p and q will be approximately constant. Those intervals in which p and q vary appreciably are excluded from consideration. In solving Problem 4 of Section 1 the following results were obtained for equations with constant coefficients:

1. If q is positive and greater than $\frac{1}{4}p^2$, the solutions are oscillatory with decreasing or increasing amplitude according as p is positive or negative; otherwise the solution is nonoscillatory.
2. The frequency and period of the oscillations are

$$\sqrt{q - \tfrac{1}{4}p^2}/2\pi \qquad \text{and} \qquad 2\pi/\sqrt{q - \tfrac{1}{4}p^2}.$$

Qualitatively the same conclusions apply to equations with variable coefficients. Of course, if the period of oscillations is large compared with the interval under consideration, the oscillatory and nonoscillatory solutions will look alike.

The transformation of Section 7 reduces p to zero and, for positive q, brings into evidence the major factor affecting the amplitude of the oscillations whether or not p is constant. The new "q" is generally variable. In Section 8 the new independent variable φ is the phase of an oscillation with "variable frequency" $\sqrt{q}$. Equation (49) shows, however, that oscillations of variable frequency are accompanied by an amplitude variation depending on the relative change in frequency per unit change in phase.

For more detail on this subject the reader is referred to E. L. Ince, *Ordinary Differential Equations*, Longmans, Green and Company, London, 1927, Chapter X (on the Sturm-Liouville theory of linear differential equations).

10. *Linear systems of equations*

Linear equations of the second order occur in the theory of electrical and mechanical oscillations of simple systems, with time as the independent

variable. Usually the parameters of the system are constant and the equations have constant coefficients; but in electrical systems, at least, it is easy to make the inductance, capacitance and resistance variable with time, in which case we have a general linear equation.

Another important application is found in the theory of the propagation of waves, electrical, mechanical, sound, etc. There, the equations appear first as systems of linear equations of the first order with two independent variables. In an electrical transmission line the voltage V across the line and the current I in it obey the following equations*:

$$\frac{dV}{dx} = Z(x)I, \qquad \frac{dI}{dx} = Y(x)V, \tag{51}$$

where x represents the distance along the line, $Z(x)$ the series impedance per unit length, and $Y(x)$ the shunt susceptance per unit length. Similar equations connect the force and velocity in a string under tension or in a spring; the coefficient Z depends on the mass per unit length and Y on the tension of the string or stiffness of the spring.

Eliminating either I or V, two second-order equations are obtained

$$\frac{d}{dx}\left(\frac{1}{Z}\frac{dV}{dx}\right) = YV, \qquad \frac{d}{dx}\left(\frac{1}{Y}\frac{dI}{dx}\right) = ZI, \tag{52}$$

or in an expanded form

$$\frac{d^2V}{dx^2} - \frac{Z'}{Z}\frac{dV}{dx} - YZV = 0, \qquad \frac{d^2I}{dx^2} - \frac{Y'}{Y}\frac{dI}{dx} - YZI = 0. \tag{53}$$

Conversely any second-order equation can be reduced to a linear system of the form (51). Thus

$$\begin{aligned} Z'/Z &= -p(x), \qquad Z(x) = \exp\left[-\int^x p(x)\,dx\right]; \\ YZ &= -q(x), \qquad Y = -q(x)\exp\left[\int^x p(x)\,dx\right]. \end{aligned} \tag{54}$$

The linear system (51) and the contracted forms (52) of the second-order equations play an extremely important role both in applications and in mathematical theory.

Let us introduce the following two functions

$$K(x) = \sqrt{Z(x)/Y(x)}, \qquad \Gamma(x) = \sqrt{Z(x)Y(x)} \tag{55}$$

* In the usual form these equations appear with negative signs in front of Z and Y. A reversal in the sign of x makes these signs positive.

into (51); then

$$\frac{dV}{dx} = K\Gamma I, \qquad \frac{dI}{dx} = K^{-1}\Gamma V. \tag{56}$$

A symmetrical form is obtained if these equations are divided by Γ and a new independent variable is defined as

$$u = \int^x \Gamma(x)\, dx. \tag{57}$$

The system becomes

$$\frac{dV}{du} = KI, \qquad \frac{dI}{du} = K^{-1}V. \tag{58}$$

Eliminating either I or V, we have

$$\frac{d^2V}{du^2} - \frac{K'}{K}\frac{dV}{du} - V = 0, \qquad \frac{d^2I}{du^2} + \frac{K'}{K}\frac{dI}{du} - I = 0. \tag{59}$$

If K is constant, the solutions are

$$V = A \exp\left[\pm\int^x \Gamma(x)\, dx\right], \qquad I = B \exp\left[\pm\int^x \Gamma(x)\, dx\right]. \tag{60}$$

11. *Liouville's approximation*

If $K(u)$ of the preceding section is not constant, we remove the first derivative by the method of Section 7. Since $K'/K = D_u(\log K)$, the transformation is

$$V = \sqrt{K}\,\hat{V}, \tag{61}$$

and the equation for $\hat{V}$ becomes

$$\frac{d^2\hat{V}}{du^2} = \left[1 + \frac{3(K')^2}{4K^2} - \frac{K''}{2K}\right]\hat{V}. \tag{62}$$

If K' and K'' are small relative to K, then the approximate solution of (51) and (52) is

$$V(x) \simeq A\sqrt[4]{Z/Y}\exp\left(\pm\int^x \sqrt{ZY}\, dx\right). \tag{63}$$

This approximation was obtained by Liouville in 1837 in a somewhat less general form. Liouville started with

$$y'' = -f(x)y, \tag{64}$$

and arrived at the approximate solution

$$y \simeq \frac{A}{\sqrt[4]{f(x)}} \exp\left[\pm i \int^x \sqrt{f(x)}\, dx\right]. \tag{65}$$

Since his time this approximation has been forgotten and rediscovered by several applied mathematicians.

For example, the Bessel equation may be expressed as

$$\frac{d}{dx}\left(x\frac{dy}{dx}\right) = \left(\frac{n^2}{x} - x\right)y. \tag{66}$$

The variable y may be compared to V in (52) where $Z = 1/x$ and $Y = (n^2/x) - x$; hence approximate solutions of Bessel's equation are

$$y \simeq \frac{A}{\sqrt[4]{n^2 - x^2}} \exp\left[\pm \int^x \sqrt{\frac{n^2}{x^2} - 1}\, dx\right]. \tag{67}$$

As x increases,

$$y \simeq A'\, x^{-1/2} e^{\pm ix}. \tag{68}$$

Considering the simplicity and generality of Liouville's approximation it would be difficult to overestimate its importance in applied mathematics.

Problems

1. Show that if we set $I = K^{-1/2}\hat{I}$ in (59), then

$$\frac{d^2\hat{I}}{du^2} = \left[1 - \frac{(K')^2}{4K^2} + \frac{K''}{2K}\right]\hat{I}.$$

2. Show that, if $u = i\varphi$, the preceding equation becomes

$$\frac{d^2\hat{I}}{d\varphi^2} = \left[-1 - \frac{(K')^2}{4K^2} + \frac{K''}{2K}\right]\hat{I},$$

where the primes now denote differentiation with respect to φ.

3. Show that, if $u = i\varphi$, (62) becomes

$$\frac{d^2\hat{V}}{d\varphi^2} = \left[-1 + \frac{3}{4K^2}\left(\frac{dK}{d\varphi}\right)^2 - \frac{1}{2K}\frac{d^2K}{d\varphi^2}\right]\hat{V}.$$

4. Obtain Liouville's approximations to the solutions of the equations in the problems in Section 8.

5. Show that the substitutions $\varphi = \int^x \sqrt{f(x)}\, dx$ and

$y(x) = [f(x)]^{-1/4}v(\varphi)$ in equation (64) lead to

$$\frac{d^2v}{d\varphi^2} = -\left[1 - \frac{1}{16}\left(\frac{1}{f}\frac{df}{d\varphi}\right)^2 - \frac{1}{4}\frac{d}{d\varphi}\left(\frac{1}{f}\frac{df}{d\varphi}\right)\right]v$$
$$= -\left[1 + \frac{5}{16f^3}\left(\frac{df}{dx}\right)^2 - \frac{1}{4f^2}\frac{d^2f}{dx^2}\right]v.$$

Liouville's approximation (65) will be good if the bracketed term differs but little from unity.

12. *Physical interpretation of Liouville's approximation*

If the parameters Z, Y of an electrical transmission line are constant, the waves may be expressed in terms of two progressive waves moving in opposite directions. The propagation constant is the square root of the product of these parameters. In a semi-infinite transmission line or in a line terminated into its characteristic impedance $K = \sqrt{Z/Y}$, only one progressive wave is present — the one moving away from the generator. The other wave appears as a reflection caused by an impedance mismatch. We should expect, therefore, that even if Z, Y are varying, but in such a way that the characteristic impedance K remains constant, there will be no reflected wave and that strictly progressive waves are possible. Indeed, we have seen that in this case exact solutions (60) of this form are possible.

If K is slowly varying, we might in the first approximation neglect reflections and consider the line as continuously matched. Since the exact match presupposes transformers distributed along the line, the voltage should increase as the square root of the characteristic impedance; this is precisely what we have in (63). The current should decrease as the square root of K and the corresponding Liouville's approximation contains the factor $(Y/Z)^{1/4}$.

Liouville's approximation is good only when K' and K'' are small compared with K. When this condition is not satisfied we rely on the "wave perturbation method" described in the next section.

13. *The wave perturbation method*

Solutions of a linear equation with variable coefficients may be regarded as distorted solutions of an equation with constant coefficients. The problem is to find some method of calculating this distortion or *perturbation*.

One method is suggested by the idea itself. Suppose that the equation to be solved is (64) and suppose for the present that $f(x)$ is positive in the interval under consideration. Let β^2 be *some* mean value of $f(x)$ in the interval. Let us rewrite (64) as follows

$$y'' = -\beta^2 y + [\beta^2 - f(x)]y, \tag{69}$$

assume that we know the last term, regard the equation as a nonhomogeneous equation with constant coefficients, and use (29). Thus we obtain

$$y(x) = y_0(x) + \frac{1}{\beta}\int_a^x [\beta^2 - f(\xi)]y(\xi) \sin \beta(x - \xi)\, d\,\xi, \tag{70}$$

where $y_0(x)$ is *any* solution of

$$y_0''(x) = -\beta^2 y_0(x). \tag{71}$$

In particular, we can choose (27) where $y_0(x)$ and its derivative assume given values at $x = a$; these values are also the initial values of $y(x)$ and its derivative. If the difference $\beta^2 - f(x)$ approaches zero as x increases to infinity, then $y_0(x)$ may conveniently be chosen as $A \cos \beta x + B \sin \beta x$ or as $A \exp i\beta x + B \exp (-i\beta x)$, where A and B are left arbitrary.

As the first approximation to $y(x)$ we take $y_0(x)$; substituting $y_0(x)$ in the integral and evaluating, we obtain the first perturbation $y_1(x)$; in the same way we obtain the second perturbation $y_2(x)$; etc. To summarize:

$$y(x) = y_0(x) + y_1(x) + y_2(x) + \cdots + y_n(x) + \cdots, \tag{72}$$

where

$$y_n(x) = \frac{1}{\beta}\int_a^x [\beta^2 - f(\xi)]\, y_{n-1}(\xi) \sin \beta(x - \xi)\, d\xi. \tag{73}$$

If $f(\xi)$ is bounded in the given interval, the series is convergent.

Consider for example the Bessel equation of order zero ($n = 0$)

$$xy'' + y' + xy = 0, \qquad y'' + \frac{1}{x}y' + y = 0. \tag{74}$$

Remove the term containing the first derivative

$$y = x^{-1/2}v, \qquad v'' = -\left(1 + \frac{1}{4x^2}\right)v. \tag{75}$$

If x is not too small (greater than unity, let us say), a convenient choice is $\beta = 1$ and $a = \infty$ when (70) becomes

$$v(x) = v_0(x) - \frac{1}{4}\int_\infty^x \frac{\sin\,(x - \xi)}{\xi^2}\, v(\xi)\, d\xi. \tag{76}$$

Choosing

$$v_0(x) = e^{ix}, \tag{77}$$

we obtain

$$v_1(x) = -\frac{1}{4}\int_\infty^x \frac{\sin(x-\xi)}{\xi^2} e^{i\xi}\,d\xi = -\frac{1}{8i}\int_\infty^x (e^{ix} - e^{-ix+2i\xi})\frac{d\xi}{\xi^2}$$
$$= \frac{1}{8i}\int_\infty^x (e^{ix} - e^{-ix+2i\xi})\,d(1/\xi). \tag{78}$$

Integrating by parts,

$$v_1(x) = \left.\frac{e^{ix} - e^{-ix+2i\xi}}{8i\xi}\right|_\infty^x + \frac{1}{4}e^{-ix}\int_\infty^x \frac{e^{2i\xi}}{\xi}\,d\xi. \tag{79}$$

The first term vanishes at both limits; the integral cannot be expressed in a closed form containing elementary functions but is of such importance in various applications that it has been given a name: the *exponential integral.* Numerical tables are usually given for the real and imaginary parts, the *cosine integral* and *sine integral.* In the conventional symbolism* (79) becomes

$$v_1(x) = \tfrac{1}{4}\left[\mathrm{Ci}\,2x + i\left(\mathrm{Si}\,2x - \frac{\pi}{2}\right)\right]e^{-ix}. \tag{80}$$

Stopping with this perturbation, we have

$$y(x) \simeq x^{-1/2}\left\{e^{ix} + \tfrac{1}{4}\left[\mathrm{Ci}\,2x + i\left(\mathrm{Si}\,2x - \frac{\pi}{2}\right)\right]e^{-ix}\right\}. \tag{81}$$

Since the coefficients in (74) are real, the real and imaginary parts of (81) are also solutions. Similarly the conjugate of (81) is a solution. In physical applications the coefficient of $\exp(-ix)$ would be called the reflection coefficient for waves originating at $x = \infty$ and coming toward $x = 0$.

The second perturbation, $y_2(x)$, would have to be computed by numerical integration. However, even if x is comparable to unity, (81) is a fair approximation to the solution.

If x is much smaller than unity, the above approximation is bound to be bad; for then $v/4x^2$ in (75) is larger than v, the perturbation from the sinusoidal solution would have to be large and many terms of the series (72) would be needed. In this region it is better to begin with a different first approximation. If x is small, the transformation (75) introduces higher powers of the large quantity, $1/x$, and it is best to deal directly with the original equation (74). Let us write (74) in the form

$$\frac{d}{dx}\left(x\frac{dy}{dx}\right) = -xy. \tag{82}$$

* See the Chapter on Sine and Cosine Integrals.

As x approaches zero, the right side will approach zero provided y does not become infinite too rapidly. Then, for small values of x, (82) may be regarded as a perturbation of a simpler equation

$$\frac{d}{dx}\left(x\frac{dy_0}{dx}\right) = 0, \qquad x\frac{dy_0}{dx} = A, \qquad \frac{dy_0}{dx} = \frac{A}{x},$$

$$y_0 = A \log x + B. \tag{83}$$

Again let us assume that the right side in (82) is known, and regard this equation as a nonhomogeneous equation whose associated homogeneous equation is (83). In equation (22) Green's function $G(x,\xi)$ is that particular solution of the homogeneous equation — equation (83) in the present case — which vanishes at $x = \xi$ and whose derivative there is unity; hence

$$G(x,\xi) = \xi \log (x/\xi). \tag{84}$$

Remembering that (22) refers to the standard form of (74), we have

$$y(x) = A \log x + B - \int_a^x \xi y(\xi) \log (x/\xi)\, d\xi. \tag{85}$$

The stage has now been set for obtaining $y(x)$ by successive perturbations of $A \log x + B$. If a is chosen to be different from zero, then A and B may be expressed in terms of $y(a)$ and $y'(a)$; thus

$$y_0(x) = ay'(a) \log (x/a) + y(a). \tag{86}$$

We can also choose $a = 0$; then A in (85) is the limit of $xy'(x)$ as x approaches zero. A class of solutions finite at $x = 0$ is obtained by setting $A = 0$; then $B = y(0)$.

Let us return to equation (64), or its equivalent (69), and assume that we are interested in a finite interval $a \leq x \leq b$. If $f(x)$ is finite, we can define β as the square root of the average value of $f(x)$:

$$\beta = \sqrt{\frac{1}{b-a}\int_a^b f(x)\, dx}. \tag{87}$$

Another choice would be the average of the square root of $f(x)$:

$$\beta = \frac{1}{b-a}\int_a^b \sqrt{f(x)}\, dx. \tag{88}$$

The first will usually be easier to apply than the second; but in any case we should remember that β is at our disposal. The rapidity of convergence of the series of perturbations depends on the deviations of $f(x)$ from β^2.

Picard's method is a special case of the wave perturbation method.

Choosing $\beta = 0$, we have

$$y_0(x) = A + Bx = y(a) + y'(a)(x - a); \tag{89}$$

and consequently

$$y_n(x) = -\int_a^x y_{n-1}(\xi)(x - \xi)f(\xi)\,d\xi. \tag{90}$$

Picard's method has obvious limitations unless $f(x)$ is small in the chosen interval.

If $f(x)$ is negative, β is a pure imaginary; then we write $\beta = i\gamma$ and change the form of (70) into

$$y(x) = y_0(x) - \frac{1}{\gamma}\int_a^x [\gamma^2 + f(\xi)]\,y(\xi)\sinh\gamma(x - \xi)\,d\xi. \tag{91}$$

In this case equation (27) becomes

$$y_0(x) = y(a)\cosh\gamma(x - a) + \frac{1}{\gamma}y'(a)\sinh\gamma(x - a). \tag{92}$$

We can also choose

$$y_0(x) = Ae^{\gamma x} + Be^{-\gamma x}. \tag{93}$$

To summarize:

1. If the equation contains the first derivative, this derivative is removed by the method of Section 7 and the equation reduced to the form (64).

2. If $f(x)$ in the reduced equation is finite over a given finite or infinite interval, we choose β^2 as some mean value of $f(x)$ and apply (72) and (73), starting with any convenient solution of an equation with constant coefficients (71).

In the example (75) β^2 was made equal to $f(\infty)$. If the interval were finite, it would be better (from the point of view of the rapidity of convergence) to choose β in accordance with (87) or (88).

3. If $f(x)$ is small in the given interval, we might set $\beta = 0$. This value may actually be giv by (87) when $f(x)$ changes sign in the interval (a,b).

4. If $f(x)$ becomes infinite in (a,b), it is desirable and may even be necessary to perturb solutions of some equation other than equation (71). The example of the Bessel equation of order zero is an illustration.

Instead of evaluating perturbation terms of various orders we may restrict ourselves to smaller intervals in each of which we could use the purely hyperbolic or sinusoidal approximation (92). This equation may, in fact, be used successfully for interpolation and extrapolation of functions satisfying second-order homogeneous differential equations. The method is explained more fully in the Chapter on Bessel functions.

Problems

1. Show that a more accurate solution of Problem 3 in Section 8 is

$$y(x) = Ax^{-1/4}\left\{\exp\left(i\frac{2}{3}x\sqrt{x}\right) + \frac{5}{36}\left[\mathrm{Ci}\left(\frac{4}{3}x\sqrt{x}\right)\right.\right.$$

$$\left.\left. + i\left(\mathrm{Si}\left(\frac{4}{3}x\sqrt{x}\right) - \frac{\pi}{2}\right)\right]\exp\left(-i\frac{2}{3}x\sqrt{x}\right)\right\} + Bx^{-1/4}\left\{\exp\left(-i\frac{2}{3}x\sqrt{x}\right)\right.$$

$$\left. + \frac{5}{36}\left[\mathrm{Ci}\left(\frac{4}{3}x\sqrt{x}\right) - i\left(\mathrm{Si}\left(\frac{4}{3}x\sqrt{x}\right) - \frac{\pi}{2}\right)\right]\exp\left(i\frac{2}{3}x\sqrt{x}\right)\right\}$$

$$= Cx^{-1/4}\left\{\cos\left(\frac{2}{3}x\sqrt{x}\right) + \frac{5}{36}\mathrm{Ci}\left(\frac{4}{3}x\sqrt{x}\right)\cos\left(\frac{2}{3}x\sqrt{x}\right)\right.$$

$$\left. + \frac{5}{36}\left[\mathrm{Si}\left(\frac{4}{3}x\sqrt{x}\right) - \frac{\pi}{2}\right]\sin\left(\frac{2}{3}x\sqrt{x}\right)\right\}$$

$$+ Dx^{-1/4}\left\{\sin\left(\frac{2}{3}x\sqrt{x}\right) + \frac{5}{36}\left[\mathrm{Si}\left(\frac{4}{3}x\sqrt{x}\right) - \frac{\pi}{2}\right]\cos\left(\frac{2}{3}x\sqrt{x}\right)\right.$$

$$\left. - \frac{5}{36}\mathrm{Ci}\left(\frac{4}{3}x\sqrt{x}\right)\sin\left(\frac{2}{3}x\sqrt{x}\right)\right\}.$$

2. Obtain an approximate solution of $y'' = -xy$, $0 \le x \le 2$, $y(0) = 1$, $y'(0) = 0$ Use (70) with β taken from (87).

Ans. $y(x) = \cos x - \frac{1}{4}x\cos x + \frac{1}{4}(1 + 2x - x^2)\sin x$.

Note: The exact solution is $y(x) = \Gamma(\frac{2}{3})3^{-1/3}x^{1/2}J_{-1/3}(\frac{2}{3}x^{3/2})$, where Γ denotes the gamma function.

3. Obtain an approximate solution of $y'' = F(x)y$ in the vicinity of $x = a$ where $F(a) = 0$. Use (89) and (90).

Ans. $y(x) = y(a)[1 + G_2(x)] + y'(a)\{(x - a)[1 + G_2(x)] - 2G_3(x)\}$,

where $G_1(x) = \int_a^x F(\xi)\,d\xi$, $G_2(x) = \int_a^x G_1(\xi)\,d\xi$, $G_3(x) = \int_a^x G_2(\xi)\,d\xi$.

4. Obtain an approximate solution of

$$y'' = -xy, \quad -2 \le x \le 0, \quad y(0) = 1, \quad y'(0) = 0. \quad \text{Use (91).}$$

Ans. $\cosh x + \frac{1}{4}x\cosh x - \frac{1}{4}(1 + 2x + x^2)\sinh x$.

5. Transform the following system of differential equations

$$\frac{dV}{dx} = -iF(x)I, \qquad \frac{dI}{dx} = -iG(x)V, \tag{A}$$

into integral equations,

$$V(x) = V_0(x) - i\int_a^x \hat{F}(\xi)I(\xi)\cos\beta_0(x-\xi)\,d\xi - K_0\int_a^x \hat{G}(\xi)V(\xi)\sin\beta_0(x-\xi)\,d\xi,$$

$$I(x) = I_0(x) - K_0^{-1}\int_a^x \hat{F}(\xi)I(\xi)\sin\beta_0(x-\xi)\,d\xi - i\int_a^x \hat{G}(\xi)V(\xi)\cos\beta_0(x-\xi)\,d\xi, \tag{B}$$

where

$$F(x) = F_0 + \hat{F}(x), \qquad G(x) = G_0 + \hat{G}(x),$$
$$\beta_0 = \sqrt{F_0G_0}, \qquad K_0 = \sqrt{F_0/G_0}, \tag{C}$$

and $V_0(x)$, $I_0(x)$ is any pair of solutions of

$$\frac{dV_0}{dx} = -iF_0I_0, \qquad \frac{dI_0}{dx} = -iG_0V_0. \tag{D}$$

The quantities F_0, G_0 are mean values of $F(x)$ and $G(x)$ in an interval $a \le x \le b$.

In terms of the initial values

$$V_0(x) = V(a)\cos\beta_0(x-a) - iK_0I(a)\sin\beta_0(x-a),$$
$$I_0(x) = -iK_0^{-1}V(a)\sin\beta_0(x-a) + I(a)\cos\beta_0(x-a). \tag{E}$$

Another convenient choice is

$$I_0(x) = Pe^{-i\beta_0x} + Qe^{i\beta_0x}, \qquad V_0(x) = K_0Pe^{-i\beta_0x} - K_0Qe^{i\beta_0x}. \tag{F}$$

The solutions of the integral equations may then be expressed as follows:

$$V(x) = V_0(x) + V_1(x) + V_2(x) + \cdots,$$
$$I(x) = I_0(x) + I_1(x) + I_2(x) + \cdots, \tag{G}$$

where V_n, I_n are obtained by substituting V_{n-1}, I_{n-1} in the integrals in the equations (B):

$$V_n(x) = -i\int_a^x \hat{F}(\xi)I_{n-1}(\xi)\cos\beta_0(x-\xi)\,d\xi$$
$$- K_0\int_a^x \hat{G}(\xi)V_{n-1}(\xi)\sin\beta_0(x-\xi)\,d\xi,$$

$$I_n(x) = -K_0^{-1}\int_a^x \hat{F}(\xi)I_{n-1}(\xi)\sin\beta_0(x-\xi)\,d\xi$$
$$- i\int_a^x \hat{G}(\xi)V_{n-1}(\xi)\cos\beta_0(x-\xi)\,d\xi. \tag{H}$$

6. Show that corresponding to $a = 0$ in equation (E) of the preceding problem we have

$$V_1(x) = V(0)[-B(x)\cos\beta_0x + A(x)\sin\beta_0x - C(x)\sin\beta_0x]$$
$$-iK_0I(0)[A(x)\cos\beta_0x + B(x)\sin\beta_0x + C(x)\cos\beta_0x],$$

$$I_1(x) = iK_0^{-1}V(0)[B(x)\sin\beta_0 x + A(x)\cos\beta_0 x - C(x)\cos\beta_0 x]$$
$$-I(0)[A(x)\sin\beta_0 x - B(x)\cos\beta_0 x + C(x)\sin\beta_0 x],$$

where

$$A(x) = \tfrac{1}{2}\int_0^x (K_0^{-1}\hat{F} - K_0\hat{G})\cos 2\beta_0\xi\, d\xi, \qquad C(x) = \tfrac{1}{2}\int_0^x (K_0^{-1}\hat{F} + K_0\hat{G})\, d\xi,$$

$$B(x) = \tfrac{1}{2}\int_0^x (K_0^{-1}\hat{F} - K_0\hat{G})\sin 2\beta_0\xi\, d\xi.$$

7. Show that if $V_0(x) = -K_0 e^{i\beta_0 x}$, $\quad I_0(x) = e^{i\beta_0 x}$, then

$$V_1(x) = -\tfrac{1}{2}iK_0 e^{i\beta_0 x}\int_0^x [K_0^{-1}\hat{F}(\xi) + K_0\hat{G}(\xi)]\, d\xi$$
$$-\tfrac{1}{2}iK_0 e^{-i\beta_0 x}\int_0^x [K_0^{-1}\hat{F}(\xi) - K_0\hat{G}(\xi)]e^{2i\beta_0\xi}\, d\xi,)$$

$$I_1(x) = \tfrac{1}{2}ie^{i\beta_0 x}\int_0^x [K_0^{-1}\hat{F}(\xi) + K_0\hat{G}(\xi)]\, d\xi$$
$$-\tfrac{1}{2}ie^{-i\beta_0 x}\int_0^x [K_0^{-1}\hat{F}(\xi) - K_0\hat{G}(\xi)]e^{2i\beta_0\xi}\, d\xi$$

8. Introducing

$$\beta(x) = \sqrt{F(x)G(x)}, \qquad K(x) = \sqrt{F(x)/G(x)}, \tag{A}$$

in equation (A) of Problem 5 we have

$$\frac{dV}{\beta(x)\,dx} = -iK(x)I, \qquad \frac{dI}{\beta(x)\,dx} = -\frac{iV}{K(x)}. \tag{B}$$

Hence, introducing a new independent variable

$$\vartheta = \int^x \beta(x)\,dx, \qquad \frac{d\vartheta}{dx} = \beta(x), \tag{C}$$

in (B), we obtain

$$\frac{dV}{d\vartheta} = -iK(\vartheta)I, \qquad \frac{dI}{d\vartheta} = -\frac{iV}{K(\vartheta)}, \tag{D}$$

where $K(\vartheta)$ is the former $K(x)$ after the substitution from (C). Show that the following change of dependent variables,

$$V(\vartheta) = [K(\vartheta)]^{1/2}\hat{V}(\vartheta), \qquad I(\vartheta) = [K(\vartheta)]^{-1/2}\hat{I}(\vartheta), \tag{E}$$

transforms (D) into

$$\frac{d\hat{V}}{d\vartheta} = -i\hat{I} - \frac{K'}{2K}\hat{V}, \qquad \frac{d\hat{I}}{d\vartheta} = -i\hat{V} + \frac{K'}{2K}\hat{I}. \tag{F}$$

Show that approximate solutions of these equations are

$$\hat{V}(\vartheta) = -e^{i\vartheta} + e^{-i\vartheta} \int_0^\vartheta \frac{K'(\varphi)}{2K(\varphi)} e^{2i\varphi} \, d\varphi;$$
$$\hat{I}(\vartheta) = e^{i\vartheta} + e^{-i\vartheta} \int_0^\vartheta \frac{K'(\varphi)}{2K(\varphi)} e^{2i\varphi} \, d\varphi; \tag{G}$$

$$\hat{V}(\vartheta) = e^{-i\vartheta} - e^{i\vartheta} \int_0^\vartheta \frac{K'(\varphi)}{2K(\varphi)} e^{-2i\varphi} \, d\varphi;$$
$$\hat{I}(\vartheta) = e^{-i\vartheta} + e^{i\vartheta} \int_0^\vartheta \frac{K'(\varphi)}{2K(\varphi)} e^{-2i\varphi} \, d\varphi. \tag{H}$$

14. *Useful identities*

Any second-order linear equation may be expressed in the following form (see Section 10):

$$\frac{d}{dx}\left[P(x)\frac{dy}{dx}\right] = Q(x)y. \tag{94}$$

If y_1 and y_2 are two solutions of this equation, then

$$P(x)\left(y_2\frac{dy_1}{dx} - y_1\frac{dy_2}{dx}\right) = C, \tag{95}$$

$$y_2\frac{dy_1}{dx} - y_1\frac{dy_2}{dx} = \frac{C}{P(x)}, \tag{96}$$

where C is a constant.

Another identity is

$$\int_a^b Q(x)y^2\,dx = -\int_a^b P(x)\left(\frac{dy}{dx}\right)^2 dx + P(x)y\frac{dy}{dx}\Big|_a^b. \tag{97}$$

If yy' vanishes at the ends of the interval, then

$$\int_a^b Q(x)y^2\,dx = -\int_a^b P(x)\left(\frac{dy}{dx}\right)^2 dx. \tag{98}$$

To prove (95) we start with

$$\frac{d}{dx}\left[P(x)\frac{dy_1}{dx}\right] = Q(x)y_1, \qquad \frac{d}{dx}\left[P(x)\frac{dy_2}{dx}\right] = Q(x)y_2, \tag{99}$$

multiply the first equation by y_2 and the second by y_1. In the difference of the two resulting equations the left side is the derivative of the left side of (95) and the right side is zero; hence follows the identity.

For example, $y_1 = \sin x$ and $y_2 = \cos x$ are two solutions of $y'' = -y$ in which $P = -Q = 1$; (95) becomes

$$\cos^2 x + \sin^2 x = C.$$

The value of the constant is determined by the values of the solutions and their derivatives for some particular value of x. For the Bessel equation of order zero (82), $P = x$ and (96) becomes

$$y_2y_1' - y_1y_2' = C/x. \tag{100}$$

Equation (97) is obtained if (94) is multiplied by y and the left side integrated by parts. If the integration by parts is applied to the first equation in (99) after it has been multiplied by y_2, then

$$\int_a^b Qy_1y_2\,dx = Py_2\frac{dy_1}{dx}\Big|_a^b - \int_a^b P\frac{dy_1}{dx}\frac{dy_2}{dx}\,dx. \tag{101}$$

Problems

1. Transform $a(x)y'' + b(x)y' + c(x)y = 0$ into (94).

Ans. $P(x) = A\exp\left[\int^x \frac{b(x)}{a(x)}\,dx\right]$, $\quad Q(x) = -\frac{c(x)P(x)}{a(x)}$,

where A is a constant.

2. Suppose that one particular solution of (95) has been found. Find the other from (96).

Ans. $Ay_1(x)\int^x \frac{dx}{P(x)[y_1(x)]^2}$, where A is some constant.

3. Show that if $y_1(\theta)$ and $y_2(\theta)$ are two solutions of Legendre's equation (33), then $y_2y_1' - y_1y_2' = C/\sin\theta$.

4. If $n = 0$ in Legendre's equation, one solution is obviously a constant. Let this constant be unity. Using the result of Problem 2, find a second solution.
Ans. $\log\tan(\theta/2)$.

15. *Boundary value problems*

The general solution of a given second-order differential equation contains two arbitrary parameters. In any given problem these must be determined from supplementary conditions, generally known as *boundary conditions*. These conditions cannot be included in the differential equation itself.

Thus the differential equation of a string vibrating with the frequency ω radians per second is

$$\frac{d}{dx}\left[T(x)\frac{dy}{dx}\right] + \omega^2M(x)y = 0, \tag{102}$$

where T is the tension, M is the mass per unit length and y is the displacement from the neutral position. The equation is the same regardless of the conditions at the ends of the string. Similarly the equations for the

voltage V and current I in an electric transmission line are

$$\frac{d}{dx}\left[\frac{1}{L(x)}\frac{dV}{dx}\right] + \omega^2 C(x)V = 0, \qquad \frac{d}{dx}\left[\frac{1}{C(x)}\frac{dI}{dx}\right] + \omega^2 L(x)I = 0, \quad (103)$$

where L and C are respectively the series inductance and shunt capacitance per unit length. These equations are the same regardless of the conditions at the ends of the line. The line may be " open " at both ends so that the current has to vanish there; or, the line may be " shorted " at both ends so that the voltage has to vanish; or, the line may be open at one end and shorted at the other. More generally the line may be " terminated " with prescribed impedances; then the voltage-current ratios are given at the terminals.

The ends of the string and of the transmission line are the " boundaries." In two- and three-dimensional problems the boundaries are lines and surfaces. In many-dimensional problems it may be easy to find the most general solutions of the equations and very difficult to obtain the particular solutions which satisfy the required conditions at the boundaries. Thus " boundary value problems " is a good name for such problems.

In the one-dimensional case there is a great difference between the solutions obtained when the boundary conditions are imposed at one point and when they are imposed at two points. If T and M are analytic, a unique solution is obtained for any prescribed pair of values of y and dy/dx at a given point; but if we prescribe the values of y at two different points, there may be no solution or infinitely many solutions (see problems of Section 1). This is as it should be. If a string is fixed at both ends, y vanishes at both ends. Such a string vibrates freely only with certain *natural* frequencies; unless ω is one of these frequencies, (102) can have no solution. And if ω is a natural frequency, there are infinitely many solutions, the amplitude of the oscillations being arbitrary as far as the equation is concerned. If the boundary conditions are given in this case, the problem is to obtain the natural frequencies of the system. If the frequency is given, the problem is to obtain consistent boundary conditions.

Equations (102) and (103) apply to regions free from impressed forces. They occur in two classes of problems: (1) free or natural oscillations, (2) forced oscillations when the forces are applied at the ends of the string or transmission line. If the impressed forces are distributed throughout an interval, the equations become nonhomogeneous. The nonhomogeneous equation has a unique solution for every value of ω for which the associated homogeneous equation has no solution; but when the homogeneous equation *has* a solution, the nonhomogeneous equation has none.

16. *Characteristic functions*

If T and M are independent of x, the general solution of (102) is

$$y = A \cos (\omega \sqrt{M/T})x + B \sin (\omega \sqrt{M/T})x. \tag{104}$$

Suppose that the string is fixed at $x = 0,\ell$; then

$$A = 0, \qquad A \cos (\omega \sqrt{M/T})\ell + B \sin (\omega \sqrt{M/T})\ell = 0. \tag{105}$$

Since A vanishes, the second equation requires either $B = 0$ or

$$\begin{aligned} &\sin (\omega \sqrt{M/T})\ell = 0, \ (\omega \sqrt{M/T})\ell = n\pi, \ n = 0, \pm 1, \pm 2, \cdots \\ &\omega_n = n\pi/\ell \sqrt{M/T}. \end{aligned} \tag{106}$$

If $B = 0$, $y = 0$ for all x and we have a trivial case of a string at rest. If $n = 0$, the solution is also trivial. The nontrivial solutions are

$$y_n(x) = B_n \sin \omega_n x, \tag{107}$$

corresponding to the ω_n's given by (106). These solutions are called the *characteristic* or *proper functions* (also *eigenfunctions*) of the boundary value problem. The values of the parameter for which these solutions exist are called the *characteristic* or *proper values* (also *eigenvalues*). The equation satisfied by the characteristic values is called the *characteristic equation*.

For each of the following sets

$$\begin{aligned} &1) \ y(0) = y(\ell) = 0, \qquad &3) \ y(0) = y'(\ell) = 0, \\ &2) \ y'(0) = y'(\ell) = 0, \qquad &4) \ y'(0) = y(\ell) = 0, \end{aligned} \tag{108}$$

of boundary conditions there exists a set of characteristic values and functions. More generally, the boundary conditions take the following form

$$\begin{aligned} p_1 y(0) + q_1 y'(0) &= 0, \\ p_2 y(\ell) + q_2 y'(\ell) &= 0, \end{aligned} \tag{109}$$

corresponding to specified " loads " at the ends of the string.

This example is representative of the general case in which T and M are functions of x. Solutions exist only for a certain set of characteristic values of the parameter ω. For any of the boundary conditions in the set

(108), equation (97) yields

$$\omega_n^2 = \frac{\int_0^\ell T(x)(dy_n/dx)^2\,dx}{\int_0^\ell M(x)y_n^2\,dx}. \tag{110}$$

One half of the numerator of this equation is equal to the potential energy of the string when its deflection is maximum; one half of the denominator multiplied by ω_n^2 is the maximum kinetic energy; the equation itself expresses the principle of conservation of energy. Equation (110) is important primarily because it permits an approximate calculation of the characteristic values. Subsequently it will be shown that if we make a first-order error in the characteristic functions, the errors in the characteristic values obtained from (110) will be of the second order.

Since T and M are positive, the characteristic values are real.

Problems

1. Determine the characteristic values and characteristic functions for the second, third and fourth sets of the boundary conditions (108).

Ans. $A_n \cos \omega_n x, \qquad \omega_n = n\pi/\ell\sqrt{M/T};$

$B_n \sin \omega_n x, \qquad \omega_n = (2n+1)\pi/2\ell\sqrt{M/T};$

$A_n \cos \omega_n x, \qquad \omega_n = (2n+1)\pi/2\ell\sqrt{M/T}.$

2. A uniform transmission line of length ℓ is shorted at $x = 0$. Across the terminals at $x = \ell$ there is an inductance L_0; the corresponding boundary condition is $LV(\ell) + L_0V'(\ell) = 0$. Obtain the characteristic equation and characteristic functions.

Ans. $\dfrac{\tan(\omega\ell\sqrt{LC})}{\omega\ell\sqrt{LC}} = -\dfrac{L_0}{L\ell}, \qquad V_n = A_n \sin(\omega_n x\sqrt{LC}).$

17. *Orthogonality*

If $y_m(x)$ and $y_n(x)$ are characteristic functions corresponding to distinct characteristic values for any one set of boundary conditions (108), then

$$\int_0^\ell M(x)y_m y_n\,dx = 0. \tag{111}$$

Functions satisfying a relationship of this type are said to be *orthogonal.*

To prove the orthogonality we start with

$$\frac{d}{dx}\left[T(x)\frac{dy_m}{dx}\right] = -\omega_m^2 M(x)y_m,$$
$$\frac{d}{dx}\left[T(x)\frac{dy_n}{dx}\right] = -\omega_n^2 M(x)y_n, \tag{112}$$

multiply the first equation by y_n, the second by y_m, and subtract

$$y_n\frac{d}{dx}\left[T(x)\frac{dy_m}{dx}\right] - y_m\frac{d}{dx}\left[T(x)\frac{dy_n}{dx}\right] = (\omega_n^2 - \omega_m^2)M(x)y_m y_n. \tag{113}$$

The left side is an exact derivative

$$\frac{d}{dx}\left\{T(x)\left[y_n\frac{dy_m}{dx} - y_m\frac{dy_n}{dx}\right]\right\} = (\omega_n^2 - \omega_m^2)M(x)y_m y_n. \tag{114}$$

Integrating from $x = 0$ to $x = l$, we obtain (111).

The well-known relations

$$\int_0^\pi \sin mx \sin nx\, dx = 0, \qquad \text{if } m \neq n;$$
$$\int_0^\pi \cos mx \cos nx\, dx = 0, \qquad \text{if } m \neq n; \tag{115}$$

show that the sets of sines and cosines are orthogonal sets in the interval $(0,\pi)$. In the interval $(0,2\pi)$ the combined set of sines and cosines is an orthogonal set.

18. *Expansions in series of orthogonal functions*

The possibility of expanding given functions into series of orthogonal functions is suggested by physical considerations, by the fact that at a particular instant we are at liberty to assign a reasonably arbitrary shape to the string. Once we admit the possibility of the following expansion

$$f(x) = \sum_{n=1}^{\infty} A_n y_n(x), \tag{116}$$

it is easy to obtain the coefficients. We simply multiply the above equation by $M(x)y_m(x)$ and integrate from $x = 0$ to $x = l$. Since the y's are

orthogonal, all terms on the right except one vanish and

$$A_m \int_0^\ell M(x)[y_m(x)]^2\,dx = \int_0^\ell f(x)M(x)y_m(x)\,dx, \qquad (117)$$

$$A_m = \frac{\int_0^\ell f(x)M(x)y_m(x)\,dx}{\int_0^\ell M(x)[y_m(x)]^2\,dx}.$$

Since a characteristic function $y_n(x)$ may be multiplied by an arbitrary constant, we can always construct a set for wh [illegible]

$$\int_0^\ell M(x)[y_n(x)]^2\,dx \; [illegible] \qquad (118)$$

Such a set is said to be *normalized*. For a [illegible] he coefficients of (116) are

$$A_n = \int_0^\ell f(x)M(x) \; [illegible] \qquad (119)$$

Let $f(x)$ and $g(x)$ be the displacement and velocity of the string at $t = 0$.

$$y(x,0) = f(x), \qquad \dot{y}(x,0) = g(x). \qquad (120)$$

At any subsequent time we shall have

$$y(x,t) = \sum_n (a_n \cos \omega_n t + b_n \sin \omega_n t) y_n(x), \qquad (121)$$

where the summation is extended over all the natural modes of oscillation. That is, the motion of the string is assumed to consist of all possible oscillations consistent with the boundary conditions. The time factors are assumed to be general, to take care of variations in the amplitude and phase. Differentiating (121) with respect to t, we obtain the velocity

$$\dot{y}(x,t) = \sum_n \omega_n(-a_n \sin \omega_n t + b_n \cos \omega_n t) y_n(x). \qquad (122)$$

If $t = 0$,

$$f(x) = \sum_n a_n y_n(x), \qquad g(x) = \sum_n \omega_n b_n y_n(x); \qquad (123)$$

therefore

$$a_n = \int_0^\ell f(x)M(x)y_n(x)\,dx, \qquad b_n = \omega_n^{-1} \int_0^\ell g(x)M(x)y_n(x)\,dx. \qquad (124)$$

The coefficients of the expansion (116) are obtained independently of each other. If one or more terms of the series have inadvertently been

omitted, there is nothing in the rule (117) for obtaining the other coefficients to indicate the omission. This points to the necessity of proving that the series (116) does in fact represent the given function $f(x)$. If one is careful in obtaining the characteristic functions, there is no practical danger; complete sets of orthogonal functions will be determined and correct expansions obtained. Still a legitimate doubt will remain until a proof of completeness has been carried out.

19. *Liouville's solution of nonhomogeneous equations*

The solution of the nonhomogeneous equation

$$\frac{d}{dx}\left[T(x)\frac{dy}{dx}\right] + \omega^2 M(x)y = F(x) \tag{125}$$

was expressed by Liouville in the following form:

$$y(x) = \sum_n \frac{y_n(x)\int_0^l F(x)y_n(x)\,dx}{\omega^2 - \omega_n^2}. \tag{126}$$

To derive this solution let

$$\frac{F(x)}{M(x)} = \sum_n a_n y_n(x), \qquad y(x) = \sum_n b_n y_n(x), \tag{127}$$

and substitute in (125). If $y_n(x)$ is a characteristic function of the associated homogeneous equation, then the result can be simplified

$$\sum_n b_n(\omega^2 - \omega_n^2)M(x)y_n(x) = \sum_n a_n M(x)y_n(x). \tag{128}$$

The equation becomes an identity if

$$b_n = \frac{a_n}{\omega^2 - \omega_n^2}. \tag{129}$$

Expanding $f(x) = F(x)/M(x)$ in an appropriate series of normalized orthogonal functions, we have

$$a_n = \int_0^l F(x)y_n(x)\,dx. \tag{130}$$

The solution is valid if ω is not equal to any of the characteristic values. In this case, the homogeneous equation possesses no solution different from zero and (126) is the only solution.

If $F(x)$ vanishes everywhere except in the interval $(\xi - 0, \xi + 0)$, where it is infinite in such a way that

$$\int_{\xi-0}^{\xi+0} F(x)\,dx = 1, \tag{131}$$

Comparing (138) to (102), we have

$$\frac{\partial \varphi}{\partial y} = \omega^2 M(x)y, \qquad \frac{\partial \varphi}{\partial y'} = -T(x)y'; \tag{139}$$

from this we find that φ is equal to one half of the integrand in (134). This constant factor has been dropped since it does not affect the stationary property of the integral. Incidentally $I/2$ is the difference between the kinetic and potential energy of the string, oscillating with frequency ω.

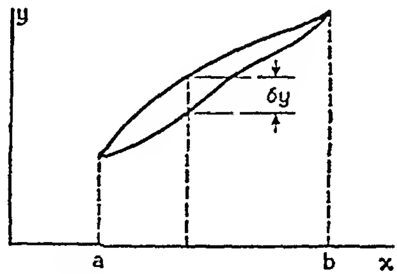

FIG. 11.1. A variation in a curve connecting two given points.

If y is one of the characteristic functions, then I is not only stationary but also vanishes (see 110). These properties can be used for the approximate evaluation of characteristic values and characteristic functions. We start with some function $y(x,A,B,C\ldots)$, depending on several parameters, which satisfies the boundary conditions. We substitute this function in (135) and determine the parameters as well as ω from

$$I = 0, \qquad \frac{\partial I}{\partial A} = 0, \qquad \frac{\partial I}{\partial B} = 0, \qquad \frac{\partial I}{\partial C} = 0, \quad \ldots . \tag{140}$$

The vanishing of the partial derivatives is required by the stationary property of I.

While considerable latitude is permitted in the choice of the approximating function, better results are obtained if this function happens to conform somewhat to the true solution. In many instances the variational method leads rapidly to simple and satisfactory approximations; but, in general, considerable numerical work is required, and the wave perturbation method of Section 13 is likely to be more powerful. In applications of that method to boundary value problems one of the initial values $y(a)$, $y'(a)$ is left arbitrary and the solution is obtained to any desired degree of accuracy; then the second boundary condition is used to determine the characteristic values of ω.

An elementary discussion of the calculus of variations is presented by William Elwood Byerly in his *Introduction to the Calculus of Variations*, Harvard University Press, Cambridge, Mass.

21. *A perturbation method for calculating the characteristic values and characteristic functions*

The wave perturbation method of Section 13 is usually well adapted to evaluating the characteristic values and characteristic functions, especially

in the case of higher modes. This method allows a subdivision of the interval $(0,\ell)$ into subintervals and the determination of the solutions of the differential equation in terms of two arbitrary constants and the parameter β. The boundary condition at one end determines one arbitrary constant. The boundary condition at the other end imposes a restriction on the parameter thus confining it to a series of characteristic values.

When the deviations of the coefficients of the differential equation from constant values are small, then the entire interval $(0,\ell)$ may be treated as a whole by the following variant of the wave perturbation method.* Let the boundary value problem be:

$$\frac{d^2y}{dx^2} + \chi^2[1 + f(x)]y = 0, \qquad y(0) = y(\ell) = 0 \tag{141}$$

where $|f(x)| \ll 1$. Let us represent $y(x)$ by the following Fourier series

$$y(x) = \sum_{n=1}^{\infty} a_n \sin (n\pi x/\ell). \tag{142}$$

The terms of this series are such that the boundary conditions are satisfied automatically. If we were to neglect $f(x)$, the various terms in (143) would be the characteristic functions corresponding to the characteristic values $\chi_n^2 = (n\pi/\ell)^2$ of the parameter χ^2; but as it is, we assume

$$\chi_n^2 = (n\pi/\ell)^2(1 + \delta_n), \tag{143}$$

and expand each characteristic function in a Fourier series,

$$Y_n(x) = \sum_{m=1}^{\infty} a_{nm} \sin (m\pi x/\ell), \qquad a_{nn} = 1. \tag{144}$$

The coefficients a_{nm} in which $m \neq n$ are small compared with unity. Substituting from (143) and (144) in (141), we obtain the following equation for the coefficients a_{nm} of $Y_n(x)$:

$$\sum_{m=1}^{\infty} \left\{-(m\pi/\ell)^2 + (n\pi/\ell)^2(1 + \delta_n)[1 + f(x)]\right\} a_{nm} \sin (m\pi x/\ell) = 0. \tag{145}$$

This equation must be satisfied for all values of x. Since δ_n and $f(x)$ are small we shall neglect their product; $(\pi/\ell)^2$ may be canceled.

To obtain δ_n and a_{nm}, we make use of the orthogonality of the sine functions in the interval $(0,\ell)$. Thus multiplying by $\sin (n\pi x/\ell)$ and integrating from 0 to ℓ, we obtain

$$\tfrac{1}{2}\ell\delta_n + \sum_{m=1}^{\infty} a_{nm} \int_0^\ell f(x) \sin (m\pi x/\ell) \sin (n\pi x/\ell)\, dx = 0. \tag{146}$$

* This variant, however, was developed before the method of Section 13.

Since $a_{nm} \ll 1$ when $n \neq m$ and $a_{nn} = 1$, we have

$$\delta_n \simeq -\frac{2}{\ell}\int_0^{\ell} f(x)\sin^2(n\pi x/\ell)\,dx. \tag{147}$$

Multiplying (145) by $\sin(k\pi x/\ell)$, where $k \neq n$, integrating, and again neglecting the products of small quantities, we find

$$\tfrac{1}{2}(n^2 - k^2)\ell a_{nk} + n^2\int_0^{\ell} f(x)\sin(n\pi x/\ell)\sin(k\pi x/\ell)\,dx = 0, \tag{148}$$

$$a_{nk} \simeq \frac{2n^2}{(k^2 - n^2)\ell}\int_0^{\ell} f(x)\sin(n\pi x/\ell)\sin(k\pi x/\ell)\,dx.$$

Equation (147) for the first-order correction to the characteristic values is particularly simple. Equations (148), however, do not lead to such convenient expressions for the characteristic functions as the wave perturbation method in Section 13. The present method fails when the coefficient of the second term in (141) is a fairly large negative quantity within the interval $(0,\ell)$; but the wave perturbation method remains applicable.

CHAPTER XII

Differential Equations of Higher Order

This brief chapter is devoted largely to differential equations with constant coefficients. The treatment is simply an extension of the method explained early in the preceding chapter. A much more powerful method is given in the Chapter on Linear Analysis.

1. *Linear homogeneous equations with constant coefficients — the characteristic equation*

The solution of the nth-order equation with constant coefficients,

$$a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_2 y'' + a_1 y' + a_0 y = 0, \tag{1}$$

is similar to the solution of the second-order equation. Since all derivatives of an exponential function are proportional to the function itself, it is clear that with a proper choice of γ the exponential function

$$y = Ae^{\gamma x} \tag{2}$$

should satisfy (1). The factor A is arbitrary. Substituting in (1), we obtain the characteristic equation for γ,

$$a_n \gamma^n + a_{n-1}\gamma^{n-1} + \cdots + a_2\gamma^2 + a_1\gamma + a_0 = 0. \tag{3}$$

An algebraic equation of the nth degree has n roots if we count multiple roots the proper number of times. If all the roots are distinct we have the general solution containing n arbitrary constants

$$y = A_1 e^{\gamma_1 x} + A_2 e^{\gamma_2 x} + \cdots + A_n e^{\gamma_n x}. \tag{4}$$

Equation (1) defines the nth-order derivative and all higher derivatives in terms of the function and its $(n - 1)$ derivatives of lower order. A possible exception would be the case in which $a_n = 0$; but then the equation would not be of the nth order. Hence, so far as (1) is concerned, the values of y and its first $(n - 1)$ derivatives are arbitrary, and the solution in any form should contain at least n arbitrary constants. In the case of a supposed solution containing more than n arbitrary constants, two cases may present themselves. Usually the extra constants would make the nth-order derivative arbitrary and (1) would not be satisfied. Occasionally it may happen that in spite of the extra constants, the derivatives of orders higher than the $(n - 1)$th are not arbitrary and are consistent with the values of the derivatives of lower order. In this case some constants

could be combined with others and the number of *effective* constants reduced. For example, if two arbitrary constants, A_1 and A_2, occur only in the sum $A_1 + A_2$, we have really only one effective constant.

If γ_1 is a multiple root, (3) becomes

$$a_n(\gamma - \gamma_1)^m(\gamma - \gamma_{m+1})(\gamma - \gamma_{m+2}) \cdots (\gamma - \gamma_n) = 0 \tag{5}$$

and there are only $n - m + 1$ distinct roots. In this case there are only $n - m + 1$ effective constants in (4) and the solution is not general. The multiple root may be regarded as the limit of m distinct roots and we may expect that the following derivatives will be solutions

$$\frac{d}{d\gamma_1} e^{\gamma_1 x}, \quad \frac{d^2}{d\gamma_1^2} e^{\gamma_1 x}, \ldots, \frac{d^{m-1}}{d\gamma_1^{m-1}} e^{\gamma_1 x}. \tag{6}$$

These derivatives are powers of x multiplied by the exponential function. Linear combinations of these solutions would serve just as well, but as a rule we use the simple set,

$$xe^{\gamma_1 x}, \quad x^2e^{\gamma_1 x}, \ldots, x^{m-1}e^{\gamma_1 x}. \tag{7}$$

That these functions satisfy (1) may be verified by substitution and making use of the following identities, obtained by differentiating the left-hand side of (3) $m - 1$ times and setting $\gamma = \gamma_1$,

$$\begin{aligned} &na_n\gamma_1^{n-1} + (n-1)a_{n-1}\gamma_1^{n-2} + \cdots + a_2\gamma_1 + a_1 = 0, \\ &n(n-1)a_n\gamma_1^{n-2} + (n-1)(n-2)a_{n-1}\gamma_1^{n-3} + \cdots + a_2 = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{8}$$

That the first $m - 1$ derivatives vanish when $\gamma = \gamma_1$ may be seen from (5).

2. *Initial conditions*

If all the roots of the characteristic equation are simple so that (4) is the general solution, the coefficients are obtained from the initial values of y and its derivatives by solving the following set of linear equations:

$$\begin{aligned} A_1e^{\gamma_1 a} + A_2e^{\gamma_2 a} + \cdots + A_ne^{\gamma_n a} &= y(a), \\ \gamma_1A_1e^{\gamma_1 a} + \gamma_2A_2e^{\gamma_2 a} + \cdots + \gamma_nA_ne^{\gamma_n a} &= y'(a), \\ \gamma_1^2A_1e^{\gamma_1 a} + \gamma_2^2A_2e^{\gamma_2 a} + \cdots + \gamma_n^2A_ne^{\gamma_n a} &= y''(a), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \gamma_1^{n-1}A_1e^{\gamma_1 a} + \gamma_2^{n-1}A_2e^{\gamma_2 a} + \cdots + \gamma_n^{n-1}A_ne^{\gamma_n a} &= y^{(n-1)}(a). \end{aligned} \tag{9}$$

A *basic set** of solutions is obtained if, successively, one of the terms on

* Any set of linearly independent solutions of a differential equation is called a *fundamental set* or *system*; hence, a basic set is always a fundamental set but not every fundamental set is a basic set. The term "basic" for the special duty, as defined in the text, was suggested by Dr. Hamming.

and find

$$y(x) = \frac{Axe^{\gamma x}}{Z'(\gamma)}, \tag{16}$$

provided $Z'(\gamma) \neq 0$.

Another simple case is

$$f(x) = A \cos \beta x = \tfrac{1}{2}A(e^{i\beta x} + e^{-i\beta x}). \tag{17}$$

If $f(x) = f_1(x) + f_2(x)$, then a solution corresponding to $f(x)$ is the sum of solutions corresponding to f_1 and f_2 separately. Therefore

$$y(x) = \tfrac{1}{2}A\left[\frac{e^{i\beta x}}{Z(i\beta)} + \frac{e^{-i\beta x}}{Z(-i\beta)}\right] = \mathrm{re}\,\frac{Ae^{i\beta x}}{Z(i\beta)}. \tag{18}$$

Similarly for $f(x) = A \sin \beta x$, we have

$$y(x) = \frac{A}{2i}\left[\frac{e^{i\beta x}}{Z(i\beta)} - \frac{e^{-i\beta x}}{Z(-i\beta)}\right]. \tag{19}$$

More generally, if $f(x) = A \cos (\beta x + \varphi)$, then

$$y(x) = \mathrm{re}\,\frac{A \exp i(\beta x + \varphi)}{Z(i\beta)}. \tag{20}$$

4. *Systems of linear equations*

An example of a system of linear equations is

$$\begin{aligned}
Z_{11}\left(\frac{d}{dt}\right) I_1 + Z_{12}\left(\frac{d}{dt}\right) I_2 + \cdots + Z_{1n}\left(\frac{d}{dt}\right) I_n &= V_1(t),\\
Z_{21}\left(\frac{d}{dt}\right) I_1 + Z_{22}\left(\frac{d}{dt}\right) I_2 + \cdots + Z_{2n}\left(\frac{d}{dt}\right) I_n &= V_2(t),\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\cdots\cdots\\
Z_{n1}\left(\frac{d}{dt}\right) I_1 + Z_{n2}\left(\frac{d}{dt}\right) I_2 + \cdots + Z_{nn}\left(\frac{d}{dt}\right) I_n &= V_n(t),
\end{aligned} \tag{21}$$

where

$$Z_{mk}\left(\frac{d}{dt}\right) I = L_{mk}\frac{dI}{dt} + R_{mk}I + \frac{1}{C_{mk}}\int^t I\,dt. \tag{22}$$

This is the set of equations for a linear electric network with n meshes; the I's are the mesh currents and the V's are the voltages impressed on the corresponding meshes. In the case of the network the system is restricted by the reciprocity theorem, requiring $Z_{mk} = Z_{km}$; but the method of solution is not affected by this property.

We could try to eliminate all dependent variables except one and reduce this system to another system in which one equation contains I_1, the next I_1 and I_2, the next I_1 and I_2 and I_3, etc.; then we could proceed with a step-by-step solution. However the method which has been employed in the case of a single equation applies equally well to a system. At first let us suppose that the system is homogeneous and assume a possible solution in the form

$$I_1 = A_1 e^{pt}, \qquad I_2 = A_2 e^{pt}, \cdots, I_n = A_n e^{pt}. \tag{23}$$

Substituting in the left-hand side of (21), we obtain a homogeneous system of linear algebraic equations in the A's,

$$\begin{aligned} Z_{11}(p)A_1 + Z_{12}(p)A_2 + \cdots + Z_{1n}(p)A_n &= 0, \\ Z_{21}(p)A_1 + Z_{22}(p)A_2 + \cdots + Z_{2n}(p)A_n &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Z_{n1}(p)A_1 + Z_{n2}(p)A_2 + \cdots + Z_{nn}(p)A_n &= 0, \end{aligned} \tag{24}$$

where

$$Z_{mk}(p) = L_{mk}p + R_{mk} + \frac{1}{C_{mk}p}. \tag{25}$$

There are only $n - 1$ effective unknowns in this system since the equations can be divided by some A_m; hence a nontrivial solution (not all the A's are equal to zero) is possible only if p is a root of the following characteristic equation,

$$D(p) = \begin{vmatrix} Z_{11}(p) & Z_{12}(p) & \cdots Z_{1n}(p) \\ Z_{21}(p) & Z_{22}(p) & \cdots Z_{2n}(p) \\ \cdots & \cdots & \cdots \\ Z_{n1}(p) & Z_{n2}(p) & \cdots Z_{nn}(p) \end{vmatrix} = 0. \tag{26}$$

The degree of this equation is $2n$. If the coefficients are real, any complex roots occur in conjugate pairs.

For each characteristic value of p the A's may be determined from (24) except for a common factor; thus the relative currents (23) in the various meshes may be determined but not the general level. Exceptions arise only when some of the roots of the characteristic equation are multiple. One way of dealing with such cases is to begin with the case of distinct roots and let some of them coalesce. Certain solutions would tend to become equal. Forming their differences and removing factors which approach zero, we obtain the solutions for the degenerate case.

If the V's in the nonhomogeneous case are exponential or circular functions, particular solutions may be obtained by assuming exponential forms

for the I's and converting the differential equations into linear algebraic equations, as explained in Chapter I. In the general case the most practical method depends on Laplace transforms.

5. *Linear equations with variable coefficients*

The wave perturbation method explained in the preceding chapter applies to equations of any order. Suppose that the equation is homogeneous; choose suitable mean values of the coefficients and replace the given equation by

$$\begin{aligned} y^{(n)} + \bar{a}_{n-1}y^{(n-1)} + \bar{a}_{n-2}y^{(n-2)} + \cdots + \bar{a}_0 y &= F(x,y,y',\cdots), \\ F(x,y,y',\cdots) &= (\bar{a}_{n-1} - a_{n-1})y^{(n-1)} + \cdots + (\bar{a}_0 - a_0)y, \end{aligned} \tag{27}$$

where the $\bar{a}$'s are mean values of the a's in a given interval; treat F as if it were a known function and write the general solution in the form

$$y(x) = y_0(x) + \int_a^x F\left(\xi, y, \frac{dy}{d\xi}, \cdots\right) G(x - \xi)\, d\xi, \tag{28}$$

where $y_0(x)$ is the general solution of the equation for which $F = 0$, and $G(x - \xi)$ is that particular solution which, together with its first $n - 2$ derivatives, vanishes at $x = \xi$. The $(n - 1)$th derivative of G should be unity at $x = \xi$.

If $y_1(x)$ is the value of the integral when $y(\xi) = y_0(\xi)$ and $y_n(x)$ is the value when $y(\xi) = y_{n-1}(\xi)$, then $y(x) = y_0(x) + y_1(x) + \cdots$.

For example, let the equation be

$$y''' + \tfrac{1}{2}xy = 0. \tag{29}$$

Consider the interval $1 \leq x \leq 3$; choose $\bar{a}_0$ equal to the average value of $a_0(= \tfrac{1}{2}x)$,

$$\bar{a}_0 = \tfrac{1}{4}\int_1^3 x\, dx = 1; \tag{30}$$

and rewrite (29) as

$$y''' + y = (1 - \tfrac{1}{2}x)y. \tag{31}$$

The characteristic equation of

$$y_0''' + y_0 = 0 \tag{32}$$

is

$$\gamma^3 + 1 = 0, \qquad \gamma_1 = -1, \qquad \gamma_{2,3} = \frac{1}{2} \pm i\frac{\sqrt{3}}{2}. \tag{33}$$

The general solution of (32) is

$$y_0 = Ae^{-x} + e^{x/2}[B\cos(x\sqrt{3}/2) + C\sin(x\sqrt{3}/2)]. \tag{34}$$

To obtain $G(x-\xi)$ it is best to start with the symmetrical form of the general solution of (32),

$$G(x-\xi) = Pe^{\gamma_1(x-\xi)} + Qe^{\gamma_2(x-\xi)} + Re^{\gamma_3(x-\xi)}. \tag{35}$$

The boundary conditions for $G(x-\xi)$ are: $G(0) = G'(0) = 0, G''(0) = 1$; therefore

$$\begin{aligned} P + Q + R &= 0,\\ \gamma_1 P + \gamma_2 Q + \gamma_3 R &= 0,\\ \gamma_1^2 P + \gamma_2^2 Q + \gamma_3^2 R &= 1. \end{aligned} \tag{36}$$

Solving,

$$P = \tfrac{1}{3}, \qquad Q = -\tfrac{1}{6}(1 + i\sqrt{3}), \qquad R = \tfrac{1}{6}(-1 + i\sqrt{3}); \tag{37}$$

hence

$$\begin{aligned} G(x-\xi) &= \tfrac{1}{3}e^{\xi-x} - \tfrac{1}{3}e^{(x-\xi)/2}\cos\frac{(x-\xi)\sqrt{3}}{2}\\ &\quad + \frac{\sqrt{3}}{3}e^{(x-\xi)/2}\sin\frac{(x-\xi)\sqrt{3}}{2}. \end{aligned} \tag{38}$$

Solutions of (29) are then obtained from

$$y(x) = y_0(x) + \int_1^x (1 - \tfrac{1}{2}\xi)y(\xi)G(x-\xi)\,d\xi \tag{39}$$

by successive integration.

CHAPTER XIII

Partial Differential Equations

1. *Partial differential equations*

A *partial differential equation* is an equation containing partial derivatives. The *order* of the equation is the order of its highest derivative; the *degree* is the degree of this derivative. The equation is *linear* if the dependent variable and its derivatives occur only in the first degree and if there are no cross-products. Many important equations of mathematical physics are linear equations and some have constant coefficients. In the case of electromagnetic equations the linearity holds for a very wide range of values of the dependent variables; but the linearity of other equations is the result of approximations based on the assumption that the dependent variables are small. The following is a list of some of the more important equations:

One-dimensional wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}. \tag{1}$$

One-dimensional diffusion equation (heat flow)

$$\frac{\partial^2 y}{\partial x^2} = k\frac{\partial y}{\partial t}. \tag{2}$$

Two-dimensional Laplace's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0. \tag{3}$$

Two-dimensional Poisson's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = f(x,y). \tag{4}$$

Telegraphist's equations

$$\frac{\partial V}{\partial x} = -RI - L\frac{\partial I}{\partial t}, \qquad \frac{\partial I}{\partial x} = -GV - C\frac{\partial V}{\partial t}. \tag{5}$$

Two-dimensional wave equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}. \tag{6}$$

Three-dimensional Poisson's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = f(x,y,z), \tag{7}$$

including Laplace's equation for which $f(x,y,z) = 0$.

Three-dimensional wave equation in dissipative media

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} + \frac{2k}{v}\frac{\partial V}{\partial t} + f(x,y,z;t). \tag{8}$$

For nondissipative media $k = 0$. For source-free media $f = 0$. This equation includes (7) as a special case when V and f are independent of time.

Maxwell's equations (generalized)

$$\begin{aligned} \operatorname{curl} E &= -\mu\frac{\partial H}{\partial t} - M(x,y,z;t), \\ \operatorname{curl} H &= gE + \epsilon\frac{\partial E}{\partial t} + J(x,y,z;t). \end{aligned} \tag{9}$$

In the original Maxwell's equations for *source-free regions*, $M = J = 0$.

Equation of transverse motion of a plate

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)^2 u = k^2\frac{\partial^2 u}{\partial t^2}. \tag{10}$$

The coefficients in all these equations are constant. Frequently, however, it is necessary to express these equations in other coordinates than cartesian; then the coefficients become variable.

It has been explained in Section 1.10 that if the source function and dependent variable are harmonic functions of time, it is convenient to regard them as the real parts of functions containing the exponential time factor $\exp(i\omega t)$. Ordinary differential equations are thus reduced to algebraic equations; in the case of partial differential equations the number of the independent variables is decreased by one. The transformation follows the pattern given in Section 1.10. Let us take equation (8), for example, and write it as

$$\frac{\partial^2 \tilde{V}}{\partial x^2} + \frac{\partial^2 \tilde{V}}{\partial y^2} + \frac{\partial^2 \tilde{V}}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 \tilde{V}}{\partial t^2} + \frac{2k}{v}\frac{\partial \tilde{V}}{\partial t} + \tilde{f}(x,y,z;t), \tag{11a}$$

to remind us that we shall be concerned with the harmonic case. We now let $\bar{V} = \text{re}\,\mathring{V} = \text{re}\,[\hat{V}\exp(i\omega t)]$, $\bar{f} = \text{re}\,\mathring{f} = \text{re}\,[\hat{f}\exp(i\omega t)]$, where $\hat{V}$ and $\hat{f}$ are independent of t. If we can find a solution of

$$\frac{\partial^2 \mathring{V}}{\partial x^2} + \frac{\partial^2 \mathring{V}}{\partial y^2} + \frac{\partial^2 \mathring{V}}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 \mathring{V}}{\partial t^2} + \frac{2k}{v}\frac{\partial \mathring{V}}{\partial t} + \mathring{f}, \tag{11b}$$

then, since the real parts of two equal complex numbers are equal, we shall have the solution of (11a). Performing the indicated differentiations with respect to t and canceling the time factor, we have the reduced equation

$$\frac{\partial^2 \hat{V}}{\partial x^2} + \frac{\partial^2 \hat{V}}{\partial y^2} + \frac{\partial^2 \hat{V}}{\partial z^2} = \sigma^2 \hat{V} + \hat{f}(x,y,z), \tag{11c}$$

where $\sigma^2 = (-\omega^2/v^2) + (2ki\omega/v)$.

2. *The problem of partial differential equations*

Partial differential equations are broad descriptions of the functions that satisfy them; most of the details are lost. All electromagnetic waves are described by Maxwell's equations; but there is a vast difference in their behavior in free space, in the presence of conducting wires, inside and outside hollow metal tubes, etc. All these differences arise from the associated *boundary conditions*. It is rarely that the most general solution of a partial differential equation is of much use; for it is likely to be unsuited to the particular solution in which we are interested — that is, to the solution which vanishes on a given surface, or which reduces to a given function at some particular instant, or which behaves in some other prescribed manner besides satisfying the equation.

A simple example will illustrate the degree of vagueness inherent in a partial differential equation. Take

$$\frac{\partial z}{\partial x} = 0, \tag{12}$$

where z is supposed to be a function of *two* variables, x and y; then

$$z = f(y), \tag{13}$$

where f is an arbitrary function. Equation (12) says merely that z does not in fact depend on x.

The general solution of the less trivial equation

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}, \tag{14}$$

is

$$z = f(x + y). \tag{15}$$

Taking the partial derivatives,

$$\frac{\partial z}{\partial x} = f'(x + y), \qquad \frac{\partial z}{\partial y} = f'(x + y), \tag{16}$$

we find that (15) is a solution of (14) regardless of the form of f (the primes indicate differentiation with respect to $x + y$).

Conversely, starting with an arbitrary function of a given function of x and y, we can obtain a relationship between the partial derivatives which holds independently of the arbitrary function. For example,

$$\begin{aligned} z = f(xy + y^2), \qquad \frac{\partial z}{\partial x} &= f'(xy + y^2)\frac{\partial}{\partial x}(xy + y^2) = yf'(xy + y^2), \\ \frac{\partial z}{\partial y} = f'(xy + y^2)\frac{\partial}{\partial y}(xy + y^2) &= (x + 2y)f'(xy + y^2), \end{aligned} \tag{17}$$

where the prime indicates the ordinary derivative with respect to $xy + y^2$, regarded as a single variable. Eliminating f', we have

$$(x + 2y)\frac{\partial z}{\partial x} = y\frac{\partial z}{\partial y}. \tag{18}$$

If we start with two arbitrary functions

$$z = f(x + y) + g(x - y), \tag{19}$$

we obtain

$$\begin{aligned} \frac{\partial z}{\partial x} &= f'(x + y) + g'(x - y), & \frac{\partial z}{\partial y} &= f'(x + y) - g'(x - y), \\ \frac{\partial^2 z}{\partial x^2} &= f''(x + y) + g''(x - y), & \frac{\partial^2 z}{\partial y^2} &= f''(x + y) + g''(x - y), \\ \frac{\partial^2 z}{\partial x \partial y} &= f''(x + y) - g''(x - y). \end{aligned} \tag{20}$$

The arbitrary functions can be eliminated from the equations in the middle line

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial^2 z}{\partial y^2}. \tag{21}$$

Let us now see if we can find a particular solution of (21) which reduces

to given functions when $x = 0$; thus the boundary conditions are

$$z = F(y), \qquad \frac{\partial z}{\partial x} = G(y), \qquad \text{at } x = 0. \tag{22}$$

Substituting $x = 0$ in the general solution (19) and in its derivative, we have

$$\begin{aligned} f(y) + g(-y) &= F(y), \\ \left.\frac{\partial z}{\partial x}\right|_{x=0} = f'(y) + g'(-y) &= G(y). \end{aligned} \tag{23}$$

Noting that $g'(-y) = dg(-y)/d(-y) = -dg(-y)/dy$, and integrating the second equation in (23) with respect to y, we obtain

$$f(y) - g(-y) = \int_a^y G(u)\,du + C. \tag{24}$$

Therefore,

$$\begin{aligned} f(y) &= \tfrac{1}{2}F(y) + \tfrac{1}{2}\int_a^y G(u)\,du + \tfrac{1}{2}C, \\ g(-y) &= \tfrac{1}{2}F(y) - \tfrac{1}{2}\int_a^y G(u)\,du - \tfrac{1}{2}C, \\ g(y) &= \tfrac{1}{2}F(-y) - \tfrac{1}{2}\int_a^{-y} G(u)\,du - \tfrac{1}{2}C. \end{aligned} \tag{25}$$

Thus we have determined the form of the arbitrary functions f, g; (19) now becomes

$$z = \tfrac{1}{2}[F(x + y) + F(y - x) + \int_{y-x}^{y+x} G(u)\,du]. \tag{26}$$

More specifically, if $F(y) = y^2$ and $G(y) = y^3$,

$$z = x^2 + y^2 + xy^3 + yx^3. \tag{27}$$

We must not be misled by our success in solving this particular boundary value problem. If we had assigned values of z on the circumference of the unit circle, we would have found ourselves in difficulties. Successful methods of solution are determined by the nature of the boundary conditions; general solutions should be sought in a form which will make it easier to satisfy the requirements at the boundaries.

The remainder of this chapter is devoted to some of the special equations mentioned in Section 1.

Problems

1. Find the general solution of

$$p^2 \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0 \qquad (p \text{ is a constant}).$$

Ans. $z = f(x + ipy) + g(x - ipy)$.

2. Find the general solution of

$$a \frac{\partial^2 z}{\partial x^2} + b \frac{\partial^2 z}{\partial x \partial y} + c \frac{\partial^2 z}{\partial y^2} = 0.$$

Ans. $z = f(x + k_1 v) + g(x + k_2 y)$, where $k_{1,2} = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2c}$.

3. Find a partial differential equation satisfied by $u = f(x + y^2)$.

Ans. $2y \dfrac{\partial u}{\partial x} = \dfrac{\partial u}{\partial y}$.

4. Eliminate the arbitrary function from $u = f(x + y^2 + xz)$.

Ans. $x \dfrac{\partial u}{\partial x} = (1 + z) \dfrac{\partial u}{\partial z}, \qquad x \dfrac{\partial u}{\partial y} = 2y \dfrac{\partial u}{\partial z}$.

3. *One-dimensional wave equation — progressive and stationary waves*

In some circumstances one-dimensional wave motion in nondissipative media is mathematically described by equation (1). The term "one-dimensional" does not imply that the actual space occupied by the waves is one-dimensional; it merely indicates that two of the three spatial coordinates are irrelevant. Waves in strings and springs, uniform plane waves of sound, uniform electromagnetic waves, are all described by equation (1). To give our ideas concrete form, let us interpret y as the transverse displacement of a string.

The general solution involving two arbitrary functions is similar to (19) since the wave equation is similar to (21); thus,

$$y(x,t) = f(x + vt) + g(x - vt). \tag{28}$$

The first term depends only on the values of $x + vt$. To an observer moving with the velocity $dx/dt = -v$, these values would appear constant; thus the first term represents a movement of a transverse displacement in the negative x direction. The second term represents another "wavelet" moving in the positive x direction. In each case there is no change in the shape and magnitude of the displacement.

Initially we are at liberty to displace various points of the string arbi-

trarily and communicate to them arbitrary velocities; thus

$$y(x,0) = F(x), \qquad \dot{y}(x,0) = G(x), \tag{29}$$

where the dot represents differentiation with respect to t. In the preceding section we have solved a problem with similar supplementary conditions; following the same method we obtain the displacement at any instant

$$y(x,t) = \tfrac{1}{2}F(x + vt) + \tfrac{1}{2}F(x - vt) + \frac{1}{2v}\int_{x-vt}^{x+vt} G(\xi)\, d\xi. \tag{30}$$

If $G = 0$, the last term vanishes; thus the initial displacement splits in half and the two halves travel in opposite directions.

In the preceding problem no restriction is imposed on the magnitude of the transverse displacement y, beyond the general assumption that it is small as required by the approximations involved in (1). If the string is held fixed at $x = 0$ and $x = l$, y must vanish there at all times. These added conditions are not satisfied by (30); but they can be satisfied if we add appropriate functions of $x - vt$ and $x + vt$.

To illustrate, let us consider a simpler situation in which a disturbance

$$y(x,t) = g(x - vt) \tag{31}$$

traveling to the left reaches a fixed end at $x = l$. To make $y(l,t) = 0$ at all times we can add $-g(x - vt)$; however, this makes $y(x,t)$ equal to zero everywhere — a trivial case. No other function of $x - vt$ will satisfy the boundary condition; but if we try a function of $x + vt$ and write

$$y(x,t) = g(x - vt) + h(x + vt), \tag{32}$$

the condition can be satisfied. Thus

$$g(l - vt) + h(l + vt) = 0, \qquad h(l + vt) = -g(l - vt). \tag{33}$$

Letting $l + vt = \xi$, we have

$$h(\xi) = -g(2l - \xi). \tag{34}$$

Substituting in (32), we find

$$y(x,t) = g(x - vt) - g(2l - x - vt). \tag{35}$$

At $x = l$, y vanishes as required; elsewhere it is unrestricted. The added term represents a wave moving in the opposite direction; except for a change in sign, it is a replica of the original wave. The formation of the second wave at $x = l$ is called *reflection;* the newly formed wave is called the *reflected wave* to distinguish it from the original or *incident* wave.

If the string is fixed at two points, reflections go on indefinitely. For example, a transverse displacement at a given point will result in the formation of two similar waves which will move in opposite directions

until they reach the ends; there they will be reflected and will move to meet each other, pass, reach the ends, etc.

A more extended disturbance can be subdivided into small displacements, and we could try to form an idea of the subsequent behavior of the string on this basis; but when there is too much overlapping between direct, reflected and re-reflected waves the picture and expressions become too complicated to be of much use.

Among special solutions of the wave equation we find

$$y(x,t) = A \cos \beta(x - vt) \text{ and } y(x,t) = A \cos \omega\left(t - \frac{x}{v}\right).$$

These solutions are essentially the same. The *profile* of the first solution at $t = 0$ is $A \cos \beta x$, which is a sine wave of wavelength $\lambda = 2\pi/\beta$; on the other hand, at each point y oscillates with a frequency $\omega = \beta v$. The second solution stresses the frequency of the wave. Either form of the solution represents a progressive sine wave moving in the positive x direction. Similarly, $A \cos \beta(x + vt)$ represents a sine wave moving in the negative x direction. The sum of the two solutions, $2A \cos \beta x \cos \omega t$, represents a *stationary wave*. The amplitude of the oscillations, $2A \cos \beta x$, varies from point to point and at some points, the *nodes*, there is no motion.

4. *One-dimensional wave equation — natural oscillations*

In the preceding section we found a solution satisfying the initial conditions first; then we modified it to satisfy the boundary conditions. Now we shall try to satisfy the boundary conditions first and then pass to the initial conditions.

The First Stage of the Solution

Let us see if the wave equation possesses solutions in the form of the product of two functions,

$$y(x,t) = X(x)T(t), \tag{36}$$

in which each factor depends on only one variable. To find the answer, we substitute in (1)

$$T\frac{d^2X}{dx^2} = \frac{1}{v^2} X \frac{d^2T}{dt^2}. \tag{37}$$

Dividing by XT, we have

$$\frac{1}{X}\frac{d^2X}{dx^2} = \frac{1}{v^2T}\frac{d^2T}{dt^2}. \tag{38}$$

The left side is independent of t and the right side is independent of x; the equation can be true only if each side is independent of both variables. Denoting by γ^2 the common value of the two sides, we have

$$\frac{d^2X}{dx^2} = \gamma^2 X, \qquad \frac{d^2T}{dt^2} = \gamma^2 v^2 T. \tag{39}$$

Thus the answer to our question is "yes"; there are infinitely many solutions of the form (36). Each factor satisfies an ordinary differential equation containing an arbitrary parameter γ. We have achieved what is known as *separation of variables*. The arbitrary constants, such as γ, introduced in the course of separation of variables, are called *separation constants*. The number of independent separation constants is one less than the number of independent variables.

This separation of variables is not always possible. For a given equation the easiest way to find out if the variables can be separated is by trial.

The Second Stage of the Solution

The next step is to introduce the boundary conditions; the separation constants will thus be restricted to a certain set of values. Suppose that at all times

$$y(0,t) = y(l,t) = 0. \tag{40}$$

Then we must have

$$X(0) = X(l) = 0. \tag{41}$$

If we impose these conditions on *every* solution of the form (36), they will be automatically imposed on *any* linear combination of such solutions.

The general solution of the first equation in (39) is

$$X = Ae^{\gamma x} + Be^{-\gamma x}. \tag{42}$$

Conditions (41) will be satisfied if

$$A + B = 0, \qquad Ae^{\gamma l} + Be^{-\gamma l} = 0. \tag{43}$$

One possible solution is the trivial one: $A = B = 0$. If A and B are not equal to zero, then

$$\frac{B}{A} = -1; \qquad \frac{B}{A} = -e^{2\gamma l}; \qquad e^{2\gamma l} = 1. \tag{44}$$

The roots of the last equation are

$$\begin{aligned} 2\gamma_n l &= 2in\pi, \qquad n = 0, \pm 1, \pm 2, \cdots \\ \gamma_n &= in\pi/l. \end{aligned} \tag{45}$$

These are the *characteristic values* of the parameter γ. Since $B = -A$ the *characteristic functions* are

$$\begin{aligned} X_n(x) &= A\,[\exp\,(in\pi x/l) - \exp\,(-in\pi x/l)] \\ &= 2iA \sin\,(n\pi x/l) = A' \sin\,(n\pi x/l). \end{aligned} \tag{46}$$

If the common value of the two sides in (38) had been denoted by $-\beta^2$ instead of γ^2, it would have been natural to write the general expression for X in the form

$$X(x) = A \cos \beta x + B \sin \beta x, \tag{47}$$

which would have led to (46) more directly. Similarly, the hyperbolic form of (42) would have been more direct. These short cuts will occur to the student as his experience increases and he begins to anticipate some of the properties of the final solution.

Substituting from (45) in the second equation in (39), we have

$$T_n(t) = C_n \cos\,(n\pi vt/l) + D_n \sin\,(n\pi vt/l). \tag{48}$$

Combining (46) with (48), we can write the most general solution of the wave equation, subject to the given boundary conditions, in the following form

$$\begin{aligned} y(x,t) &= \sum_{n=1}^{\infty} a_n X_n(x) T_n(t) \\ &= \sum_{n=1}^{\infty} [P_n \cos\,(n\pi vt/l) + Q_n \sin\,(n\pi vt/l)] \sin\,(n\pi x/l). \end{aligned} \tag{49}$$

The characteristic value $n = 0$ has been omitted since it leads to the trivial solution $y = 0$ The negative values of n lead to the same set of characteristic functions as the positive values, and their inclusion would not make (49) more general.

The Third Stage of the Solution

The third and last stage of the solution consists of calculating the unknown constants in (49) from the remaining conditions imposed on y. At this stage the problem is the one considered in Section (11.18) from a more general point of view. In that section the coefficients are determined in such a way that $y(x,t)$ and its time derivative $\dot{y}(x,t)$ reduce to given functions at the instant $t = 0$; to apply the results to the present case we should let $y_n(x) = (\sqrt{2/l}) \sin\,(n\pi x/l)$ and $M(x) = 1$. The numerical factor $(\sqrt{2/l})$ enters through the normalization of the characteristic functions.

Various characteristic functions represent different *modes of oscillation*, corresponding to different *natural frequencies*.

Problems

1. Equation (2) governs one-dimensional heat flow; y is the temperature. It may be applied to a rod of uniform cross section provided the physical conditions are such that the temperature is the same at all points of any normal cross section, and provided there is no loss of heat except through the ends of the rod. Find the general solution for the case in which the ends $x = 0$ and $x = \ell$ of the rod are kept at zero temperature.

Ans. $y(x,t) = \sum_{n=1}^{\infty} A_n \exp(-n^2\pi^2 t/k\ell^2) \sin(n\pi x/\ell)$.

2. Let the initial temperature of the rod in the preceding problem be $y(x, 0) = f(x)$. What is the temperature at a subsequent instant?

Ans. $y(x,t) = \dfrac{2}{\ell} \sum_{n=1}^{\infty} \exp(-n^2\pi^2 t/k\ell^2) \sin(n\pi x/\ell) \int_0^\ell f(x) \sin(n\pi x/\ell)\, dx$.

3. Find the general solution of Problem 1 when the ends are kept at fixed temperatures y_1 and y_2. *Hint:* Find the steady state solution in which $\partial/\partial t = 0$, and show that the difference between it and the required solution satisfies the boundary conditions of Problem 1.

Ans. $y(x,t) = y_1 + \dfrac{y_2 - y_1}{\ell} x + \bar{y}(x,t)$, where $\bar{y}(x,t)$ is given by the expansion in Problem 1.

4. Complete the solution of Problem 3 when the initial temperature distribution is $f(x)$.

Ans.
$$y(x,t) = y_1 + \frac{y_2 - y_1}{\ell} x + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-)^n y_2 - y_1}{n} \exp\left(-\frac{n^2\pi^2 t}{k\ell^2}\right) \sin\frac{n\pi x}{\ell}$$
$$+ \frac{2}{\ell} \sum_{n=1}^{\infty} \exp\left(-\frac{n^2\pi^2 t}{k\ell^2}\right) \sin\frac{n\pi x}{\ell} \int_0^\ell f(x) \sin\frac{n\pi x}{\ell}\, dx.$$

5. Consider a thin insulated rod bent into a ring. Let the initial temperature be $f(x)$; what is the subsequent distribution? *Hint:* If ℓ is the length of the ring, then at all times and places $y(x,t) = y(x + \ell, t)$.

Ans. $y(x,t) = \dfrac{1}{\ell} \displaystyle\int_0^\ell f(x)\, dx + \dfrac{2}{\ell} \sum_{n=1}^{\infty} \exp\left(-\dfrac{4n^2\pi^2 t}{k\ell^2}\right)\left(A_n \cos\dfrac{2n\pi x}{\ell} + B_n \sin\dfrac{2n\pi x}{\ell}\right)$,

where

$$A_n = \frac{2}{\ell} \int_0^\ell f(x) \cos\frac{2n\pi x}{\ell}\, dx,\quad B_n = \frac{2}{\ell} \int_0^\ell f(x) \sin\frac{2n\pi x}{\ell}\, dx.$$

6. Under what conditions do the telegraphist's equations become identical with the equations of one-dimensional heat flow?

Ans. 1) $L = G = 0$; 2) $R = C = 0$.

7. In the case of a noninductive cable ($L = 0$) V and I satisfy the same equation

$$\frac{\partial^2 V}{\partial x^2} = RGV + RC\frac{\partial V}{\partial t}.$$

This is also the equation of heat flow with radiation.
Separate the variables.

Ans. $V = X(x)T(t),\ \dfrac{d^2X}{dx^2} = -\beta^2 X,\ \dfrac{dT}{dt} = -\left(\dfrac{G}{C} + \dfrac{\beta^2}{RC}\right)T.$

8. What is the principal difference between the solutions of the equations in Problems 1 and 7?

Ans. The former are multiplied by $\exp(-Gt/C)$ to obtain the latter.

9. Since the coefficients in the equation in Problem 7 are constants, there exist solutions of the form $\exp(\gamma x + pt)$. Show that $\exp[\gamma x + (\gamma^2 t/RC) - (Gt/C)]$ is a solution for any value of γ.

5. *Two-dimensional Laplace's equation — the general solution*

Laplace's equation (3) is obtained if y in (21) is replaced by iy; hence its general solution is

$$V(x,y) = f(x + iy) + g(x - iy). \tag{50}$$

In particular

$$\begin{aligned} V(x,y) &= \tfrac{1}{2}[f(x + iy) + f(x - iy)] = \operatorname{re} f(x + iy), \\ V(x,y) &= \frac{1}{2i}[f(x + iy) - f(x - iy)] = \operatorname{im} f(x - iy), \end{aligned} \tag{51}$$

are among the solutions. Thus all analytic functions of complex variables $x + iy$, $x - iy$ as well as their real and imaginary parts are solutions of the two-dimensional Laplace's equation.

Applications of this result to boundary value problems are considered in the Chapter on Conformal Mapping.

6. *Laplace's equation — the method of characteristic functions*

The variables in the two-dimensional Laplace's equation (3) are separable and we can find particular solutions in the product form

$$V(x,y) = X(x)Y(y). \tag{52}$$

Thus substituting in (3) and dividing by XY, we obtain

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} = 0. \tag{53}$$

The sum of functions of different variables cannot vanish unless these functions degenerate into constants; therefore

$$\frac{d^2X}{dx^2} = \gamma^2 X, \qquad \frac{d^2Y}{dy^2} = -\gamma^2 Y. \tag{54}$$

General solutions of these equations are

$$X(x) = Ae^{\gamma x} + Be^{-\gamma x}, \qquad Y(y) = C\cos\gamma y + D\sin\gamma y. \tag{55}$$

Such solutions are adapted to the solution of problems involving prescribed conditions on boundaries x = constant and y = constant. Consider for instance two perfectly conducting parallel planes $y = 0, b$, Figure 13.1. If V is the electric potential, then it should be constant on the boundaries. The product XY can be constant on $y = 0, b$ only under one of the following conditions:

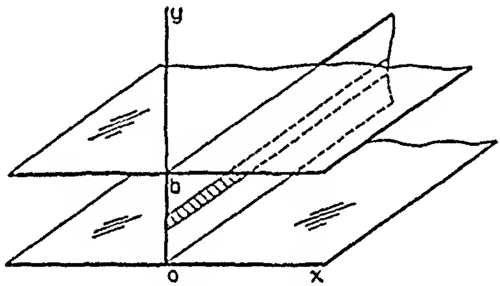

FIG. 13.1. Two parallel planes and a uniformly charged strip (indicated by the shading).

1. $X(x)$ is a constant;
2. $Y(y)$ vanishes on the boundaries

$$Y(0) = Y(b) = 0. \tag{56}$$

In the second case we must have

$$\begin{aligned} C &= 0, \qquad \sin\gamma b = 0, \\ \gamma b &= n\pi, \qquad n = \pm 1, \pm 2, \pm 3, \cdots. \end{aligned} \tag{57}$$

We exclude $n = 0$ since it yields the trivial solution $V = 0$ for all x and y. Without loss of generality we may assume that n is positive, for negative

values of n in conjunction with arbitrary constants lead to the same set of characteristic functions as the positive values. These functions are

$$V_n(x,y) = [A_n \exp(n\pi x/b) + B_n \exp(-n\pi x/b)] \sin(n\pi y/b). \quad (58)$$

Next, let us consider the case in which $X(x)$ is a constant and hence $\gamma = 0$. A direct substitution in (55) does not yield the most general solutions for this case; two of the arbitrary constants disappear altogether. Going back to (54), we find

$$\begin{aligned} X_0(x) &= A_0 x + B_0, \qquad Y_0(y) = C_0 y + D_0; \\ V_0(x,y) &= (A_0 x + B_0)(C_0 y + D_0). \end{aligned} \quad (59)$$

If V_0 is to be constant when $y = 0, b$, we must have either $A_0 = 0$ or $Y_0(0) = Y_0(b) = 0$. In the latter case Y_0 and V_0 become identically equal to zero; in the former case C_0 and D_0 are unrestricted and

$$V_0(x,y) = C_0 y + D_0. \quad (60)$$

Hence the most general form of the potential between the planes is

$$\begin{aligned} V(x,y) = A_0 y + B_0 &+ \sum_{n=1}^{\infty} A_n \exp(n\pi x/b) \sin(n\pi y/b) \\ &+ \sum_{n=1}^{\infty} B_n \exp(-n\pi x/b) \sin(n\pi y/b). \end{aligned} \quad (61)$$

The potentials of the planes determine A_0 and B_0; the infinite series allow for all kinds of variation in the potential between the planes without affecting the potentials of the planes.

The electric intensity of the field is obtained by differentiation,

$$E_x = -\frac{\partial V}{\partial x}, \qquad E_y = -\frac{\partial V}{\partial y}. \quad (62)$$

The first term in (61) yields a uniform electric field normal to the planes and implies equal and opposite charges per unit area on each plane. If both planes are at zero potential, then $A_0 = B_0 = 0$. If the field is produced by some charge distribution in the yz plane, it must not increase as we move away to infinity in either direction. In this case every A_n should be made equal to zero on the positive side of the charge distribution (where $x > 0$) and B_n should be set equal to zero on the negative side. For a sufficiently large x, only the first terms would be important

$$\begin{aligned} V(x,y) &= B_1 \exp(-\pi x/b) \sin(\pi y/b), \qquad x \to \infty; \\ &= A_1 \exp(\pi x/b) \sin(\pi y/b), \qquad x \to -\infty; \end{aligned} \quad (63)$$

and the field would eventually decrease exponentially with increasing distance from the source. Thus the potential distribution between the planes would gradually become sinusoidal.

If the source is a uniformly charged strip of width s, between $y = h$ and $y = h + s$, in the plane $x = 0$, and if the charge density is q, the following conditions must be satisfied:

$$V \text{ is continuous at } x = 0,$$

$$\frac{\partial V}{\partial x} \text{ is continuous at } x = 0 \text{ if } y < h \text{ or } y > h + s, \tag{64}$$

$$\left.\frac{\partial V(x,y)}{\partial x}\right|_{x=+0} - \left.\frac{\partial V(x,y)}{\partial x}\right|_{x=-0} = -\frac{q}{\epsilon},$$

where ϵ is the dielectric constant of the medium between the planes. The latter two conditions may be combined into one:

$$\begin{aligned} \delta = \left.\frac{\partial V(x,y)}{\partial x}\right|_{x=+0} - \left.\frac{\partial V(x,y)}{\partial x}\right|_{x=-0} &= 0 \text{ if } y < h \text{ or } y > h + s, \\ &= -q/\epsilon \text{ if } h < y < h + s. \end{aligned} \tag{65}$$

These conditions are sufficient to determine the coefficients in (61). Remembering that for $x > 0$ only the B's are present and for $x < 0$ only the A's, we have

$$\delta = -\sum_{n=1}^{\infty} \frac{n\pi}{b} (A_n + B_n) \sin (n\pi y/b). \tag{66}$$

The continuity of the potential requires $A_n = B_n$; therefore

$$\delta = -\frac{2\pi}{b} \sum_{n=1}^{\infty} nA_n \sin (n\pi y/b). \tag{67}$$

Multiplying by $\sin (m\pi y/b)$ and integrating from $y = 0$ to $y = b$, we have

$$A_m = -\frac{1}{m\pi} \int_0^b \delta \sin (m\pi y/b)\, dy. \tag{68}$$

Substituting from (65), we have

$$A_m = \frac{q}{m\pi\epsilon} \int_h^{h+s} \sin (m\pi y/b)\, dy = \frac{2bq}{m^2\pi^2\epsilon} \sin \frac{m\pi s}{2b} \sin \frac{m\pi}{b}\left(h + \frac{s}{2}\right) \tag{69}$$

Thus we have determined the coefficients of (61). The potential due to the strip of electric charge between two infinite parallel planes at zero

potential is

$$
\begin{aligned}
V(x,y) &= \sum_{n=1}^{\infty} A_n \exp(-n\pi x/b) \sin(n\pi y/b), \qquad x > 0; \\
&= \sum_{n=1}^{\infty} A_n \exp(n\pi x/b) \sin(n\pi y/b). \qquad x < 0.
\end{aligned} \tag{70}
$$

As s approaches zero, the coefficients become

$$A_n = (qs/n\pi\epsilon) \sin(n\pi h/b). \tag{71}$$

With these coefficients (70) becomes the Green's function for the present boundary value problem. The elementary source is in the plane $x = 0$; if we replace x by $x - x_0$ and h by y_0, we shall have the general expression for the Green's function representing the potential at a point (x,y) due to a source at (x_0,y_0). The field of any given charge distribution (with the density depending only on two coordinates, x and y) may then be expressed either as a sum, or a line integral, or an area integral, or as a combination of these, depending on whether the charge is concentrated in filaments parallel to the z-axis, or on cylindrical surfaces with the generators parallel to the z-axis, or distributed in volumes enclosed by such surfaces.

In the vicinity of a linear source the series in (70) converge very slowly; this feature is common to all expansions of this type. In some instances, such as the present one, there exist solutions which are equally good at all points, but in general it may be necessary to look for two or more different forms of the solution to cover the range of interest.

Problems

1. Consider a slab of conducting substance between $y = 0$ and $y = b$. The potential of the two-dimensional flow satisfies (3); the electric current density is $J = gE$, where g is the conductivity and $E = -\text{grad } V$; the normal component of the current density, and therefore the normal derivative of V, vanishes at the boundaries of the slab. Find the Green's function; that is, the potential of a line source at (x_0,y_0) emitting one ampere per unit length.

Ans.
$$
\begin{aligned}
V(x,y;x_0y_0) &= \frac{x_0 - x}{2gb} + \frac{1}{\pi g}\sum_{n=1}^{\infty} \frac{1}{n} \exp\frac{n\pi(x_0 - x)}{b} \cos\frac{n\pi y_0}{b} \cos\frac{n\pi y}{b}, \\
&\qquad x > x_0; \\
&= \frac{x - x_0}{2gb} + \frac{1}{\pi g}\sum_{n=1}^{\infty} \frac{1}{n} \exp\frac{n\pi(x - x_0)}{b} \cos\frac{n\pi y_0}{b} \cos\frac{n\pi y}{b}, \\
&\qquad x < x_0.
\end{aligned}
$$

2. Solve the preceding problem for the case in which one boundary, $y = 0$, is a

perfect conductor. At this boundary $\partial V/\partial x$ should vanish.

Ans. $$V(x,y;x_0y_0) = \frac{2}{\pi g}\sum_{n=0}^{\infty}\frac{1}{2n+1}\exp\frac{(2n+1)\pi(x_0-x)}{2b}\sin\frac{(2n+1)\pi y_0}{2b}\times \sin\frac{(2n+1)\pi y}{2b},\ x > x_0;$$

$$= \frac{2}{\pi g}\sum_{n=0}^{\infty}\frac{1}{2n+1}\exp\frac{(2n+1)\pi(x-x_0)}{2b}\sin\frac{(2n+1)\pi y_0}{2b}\times \sin\frac{(2n+1)\pi y}{2b},\ x < x_0.$$

3. Suppose that the slabs in the preceding problems are cut in two along the plane $x = 0$. Find the potentials.

Ans. In each case there is an additional series representing the potential of a unit source at $(-x_0,y_0)$.

4. Suppose that the boundary $x = 0$ in Problem 3 is made perfectly conducting. Find the potentials.

Ans. The additional series represent the potentials of a negative unit source at $(-x_0,y_0)$.

7. *Laplace's equation in cylindrical coordinates*

Let us now consider problems involving two coaxial cylinders, Figure 13.2. Solutions in the form (55) are not suitable for the new boundaries; for the success of the present method the boundaries should coincide with the coordinate surfaces and we should turn to cylindrical coordinates.

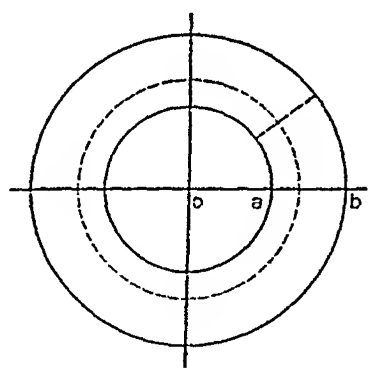

FIG. 13.2. A cross section of two coaxial cylinders. Coaxial circles and radii (dotted lines) may be used to subdivide the region into subregions in which the boundary conditions of a given problem may be satisfied.

In these coordinates Laplace's equation becomes

$$\rho^2\frac{\partial^2 V}{\partial\rho^2} + \rho\frac{\partial V}{\partial\rho} + \frac{\partial^2 V}{\partial\varphi^2} = 0. \tag{72}$$

Assuming solutions in the product form

$$V(\rho,\varphi) = R(\rho)\Phi(\varphi), \tag{73}$$

we substitute in (72) and divide by V

$$\frac{1}{R}\left(\rho^2\frac{d^2R}{d\rho^2} + \rho\frac{dR}{d\rho}\right) + \frac{1}{\Phi}\frac{d^2\Phi}{d\varphi^2} = 0. \tag{74}$$

The variables are now separated and we obtain two ordinary differential equations

$$\frac{d^2\Phi}{d\varphi^2} = -n^2\Phi, \qquad \rho^2\frac{d^2R}{d\rho^2} + \rho\frac{dR}{d\rho} - n^2R = 0, \tag{75}$$

with an arbitrary parameter n. The form of the separation constant has been chosen to make it natural to express the Φ-function in terms of circular functions

$$\Phi_n(\varphi) = A_n \cos n\varphi + B_n \sin n\varphi, \qquad R_n(\rho) = C_n\rho^n + D_n\rho^{-n}. \tag{76}$$

We should not lose sight, however, of the essential arbitrariness of n, until we begin to impose the boundary conditions. It is also important to keep in mind that to some extent we are free to choose, from a given set, those boundary conditions which are to determine the separation constants.

The solutions (76) cease to be general if $n = 0$. In this case we find directly from (75)

$$\Phi_0(\varphi) = A_0 + B_0\varphi, \qquad R_0(\rho) = C_0 + D_0 \log \rho. \tag{77}$$

We shall now solve several electrostatic problems in which the coaxial cylinders $\rho = a$ and $\rho = b$ are conductors. The simplest case is the one in which the electric charge is confined to the cylinders. On conductors the charge is free to move and our system is symmetric; hence the charge distribution and the field should be independent of φ. The only solution of this kind is

$$V(\rho) = A + B \log \rho. \tag{78}$$

If $V(a) = V_1$ and $V(b) = V_2$,

$$A + B \log a = V_1, \qquad A + B \log b = V_2;$$

$$B = \frac{V_2 - V_1}{\log (b/a)}, \qquad V(\rho) - V_1 = B \log (\rho/a); \tag{79}$$

$$V(\rho) = V_1 + \frac{V_2 - V_1}{\log (b/a)} \log (\rho/a).$$

It is not necessary to use symmetry considerations. If there is no charge between the cylinders, V should satisfy Laplace's equation for all values of φ and for all values of ρ in the interval (a,b). The potential is a single-valued function; hence in the present case it must also be periodic, with the period 2π. This condition restricts n to a set of integers. For any n different from zero, R_n should vanish when $\rho = a$ and $\rho = b$; otherwise, the associated Φ_n-factor would make the potentials of the conducting

cylinders variable. Thus

$$C_n a^n + D_n a^{-n} = 0, \qquad C_n b^n + D_n b^{-n} = 0,$$
$$-\frac{D_n}{C_n} = a^{2n} = b^{2n}, \qquad (b/a)^{2n} = 1, \tag{80}$$

or else $C_n = D_n = 0$. For integral values of n the nontrivial solution (80) is impossible and we are led back to (78).

Let us now suppose that a charged filament is passing through some point (ρ_0,φ_0). This alters the situation because Laplace's equation is no longer satisfied for *all* values of ρ and φ. However, the equation *is* satisfied,

A. For all values of φ in two regions, $a \leq \rho < \rho_0$ and $\rho_0 < \rho \leq b$;
B. For all values of ρ if $\varphi \neq \varphi_0$.

Without loss of generality we may assume that $\varphi_0 = 0$. The statement (A) applies more generally to the case in which the electric charge between the cylinders is confined to the cylinder $\rho = \rho_0$. Having found the solution for this cylindrical layer of charge, we can obtain the solution for any charge distribution by subdividing the latter into coaxial cylindrical layers. The statement (B) applies more generally to the case in which the electric charge is confined to the radial half-plane $\varphi = \varphi_0$. Having found the solution for this case, we can obtain the solution for any charge distribution by subdividing the latter into radial layers. The two methods of approach lead to different sets of characteristic functions. Any charged filament may be regarded either as a "cylindrical layer" or as a "radial layer"; thus we shall have two different series expansions for the Green's function.

We shall now study the two alternatives in greater detail. In the case (A) all Φ-functions should be periodic; therefore, $B_0 = 0$ and n is an integer. Thus we have

$$\begin{aligned} V(\rho,\varphi) &= C_0 + D_0 \log \rho + \sum_{n=1}^{\infty} (C_n\rho^n + D_n\rho^{-n})(A_n \cos n\varphi + B_n \sin n\varphi), \\ &\qquad\qquad a \leq \rho < \rho_0, \\ &= C_0' + D_0' \log \rho + \sum_{n=1}^{\infty} (C_n'\rho^n + D_n'\rho^{-n})(A_n' \cos n\varphi + B_n' \sin n\varphi), \\ &\qquad\qquad \rho_0 < \rho \leq b. \end{aligned} \tag{81}$$

We are forced to consider the two regions separately because V does not satisfy Laplace's equation when $\rho = \rho_0$. The coefficients in these expansions are found as follows. If $\rho = a$, V is constant; this makes it possible to express D_n in terms of C_n (when $n \neq 0$) or vice versa. Similarly D_n'

may be expressed in terms of C_n'. Since C_n and C_n' can be absorbed by the constants associated with the Φ-functions, we can set arbitrarily $C_n' = C_n = 1$. The potential across the charged layer $\rho = \rho_0$ should be continuous; this relates A_n''s to A_n's and B_n''s to B_n's. The radial derivative of V should be discontinuous and the magnitude of the discontinuity should be related to the charge density. For a narrow charged strip we obtain equations similar to (65). The orthogonality of the circular functions then enables us to determine the remaining coefficients. Thus, if the coaxial cylinders are kept at zero potential, for a filament at (ρ_0,φ_0) having an electric charge q per unit length we find

$$V(\rho,\varphi;\rho_0,\varphi_0) = \frac{q \log (b/\rho_0) \log (\rho/a)}{2\pi\epsilon \log (b/a)} - \frac{q}{\pi\epsilon} \sum_{n=1}^{\infty} \frac{[(\rho_0/b)^n - (b/\rho_0)^n][(\rho/a)^n - (a/\rho)^n]}{2n[(b/a)^n - (a/b)^n]} \cos n(\varphi - \varphi_0), \tag{82}$$

$$V(\rho,\varphi;\rho_0\varphi_0) = \frac{q \log (\rho_0/a) \log (b/\rho)}{2\pi\epsilon \log (b/a)} - \frac{q}{\pi\epsilon} \sum_{n=1}^{\infty} \frac{[(\rho_0/a)^n - (a/\rho_0)^n][(\rho/b)^n - (b/\rho)^n]}{2n[(b/a)^n - (a/b)^n]} \cos n(\varphi - \varphi_0), \tag{83}$$

according as $\rho < \rho_0$ or $\rho > \rho_0$. It is suggested that the reader should fill in the details and actually derive the above equations before passing on to the second case.

In the case (B) we want a single analytic expression valid for any ρ, including a and b. If V is to vanish for these extreme values we must satisfy (80). The roots of the last equation in this set are*

$$2n \log (b/a) = 2m\pi i, \qquad n = im\pi/\log (b/a), \qquad m = 1, 2, 3, \cdots. \tag{84}$$

The negative values of m are omitted because they yield essentially the same characteristic functions as the positive values. Since the coefficients of R_n are related as in (80), we have

$$\begin{aligned} R_n(\rho) &= A_n'[(\rho/a)^n - (a/\rho)^n] \\ &= A_n'\{\exp [n \log (\rho/a)] - \exp [-n \log (\rho/a)]\}. \end{aligned} \tag{85}$$

Substituting from (84) and combining the exponential functions, we obtain

$$\begin{aligned} R_m(\rho) &= A_m \sin [mk \log (\rho/a)], \qquad k = \pi/\log (b/a), \\ \Phi_m(\varphi) &= C_m e^{mk\varphi} + D_m e^{-mk\varphi}, \qquad \varphi_0 \leq \varphi \leq 2\pi + \varphi_0. \end{aligned} \tag{86}$$

* Since $x^n = \exp (n \log x)$ and $\exp (2m\pi i) = 1$ when m is an integer, we have $n \log x = 2m\pi i$.

If the potentials of the coaxial conductors are to be different from zero, we should include an additional solution of the form (78).

The Φ-functions are no longer periodic and the range of φ should be restricted to an interval $(\varphi_0, \varphi_0 + 2\pi)$. A charge distribution in the radial half-plane $\varphi = \varphi_0$ causes a discontinuity in $\partial V/\partial\varphi$ but leaves V continuous. Expressing the continuity condition, we have

$$C_m e^{mk\varphi_0} + D_m e^{-mk\varphi_0} = C_m e^{mk(\varphi_0+2\pi)} + D_m e^{-mk(\varphi_0+2\pi)},$$

$$C_m(e^{2\pi km} - 1)e^{mk\varphi_0} = D_m(1 - e^{-2\pi km})e^{-mk\varphi_0}, \tag{87}$$

$$C_m = B'_m e^{-\pi km - km\varphi_0}, \qquad D_m = B'_m e^{\pi km + km\varphi_0}.$$

Therefore

$$\begin{aligned}\Phi_m(\varphi) &= B'_m e^{mk(\varphi-\varphi_0-\pi)} + B'_m e^{-mk(\varphi-\varphi_0-\pi)} \\ &= B_m \cosh mk(\varphi - \varphi_0 - \pi).\end{aligned} \tag{88}$$

Thus the general continuous form of V is

$$V(\rho,\varphi) = \sum_{m=1}^{\infty} A_m \sin[mk \log(\rho/a)] \cosh mk(\varphi - \varphi_0 - \pi). \tag{89}$$

If $q(\rho)$ is the surface density of electric charge in the half-plane $\varphi = \varphi_0$, then in accordance with electrostatic laws

$$\frac{1}{\rho}\frac{\partial V}{\partial\varphi}\bigg|_{\varphi=\varphi_0+2\pi} - \frac{1}{\rho}\frac{\partial V}{\partial\varphi}\bigg|_{\varphi=\varphi_0} = q(\rho)/\epsilon. \tag{90}$$

Substituting from (89)

$$\frac{1}{\rho}\sum_{m=1}^{\infty} 2mkA_m \sinh mk\pi \sin[mk \log(\rho/a)] = q(\rho)/\epsilon. \tag{91}$$

This equation should be satisfied for all values of ρ.

The R-functions are solutions of

$$\frac{d}{d\rho}\left(\rho\frac{dR}{d\rho}\right) = \frac{n^2}{\rho}R. \tag{92}$$

By (11–111) the R-functions are orthogonal and the "weight-factor" $M(\rho) = 1/\rho$. Hence, multiplying (91) by $\sin[pk \log(\rho/a)]$, where p is an integer, and integrating from $\rho = a$ to $\rho = b$, we obtain

$$\begin{aligned}A_p &= \frac{\displaystyle\int_a^b q(\rho)\sin[pk\log(\rho/a)]\,d\rho}{\displaystyle 2pk\epsilon\sinh pk\pi\int_a^b \rho^{-1}\sin^2[pk\log(\rho/a)]\,d\rho} \\ &= \frac{1}{\pi p\epsilon}\operatorname{csch} pk\pi\int_a^b q(\rho)\sin[pk\log(\rho/a)]\,d\rho.\end{aligned} \tag{93}$$

If the charge is concentrated on a strip of width s, extending from $\rho = \rho_0 - \frac{1}{2}s$ to $\rho = \rho_0 + \frac{1}{2}s$, and if we let $q(\rho)$ increase as s decreases in such a way that its integral over this interval is a constant q_0,

$$A_p = \frac{q_0}{\pi p \epsilon} \operatorname{csch} pk\pi \sin [pk \log (\rho_0/a)]. \tag{94}$$

Thus the coefficients in (89) have been determined, and we have another form of the solution of our problem, essentially distinct from that given by (82) and (83). The new expansion is particularly good when b/a is near unity and φ is not too close to φ_0 or $\varphi_0 + 2\pi$.

Problems

1. Show that, as a increases indefinitely while $b - a$ remains finite, equations (89) and (94) approach (70) and (71).

2. What happens to (82) and (83) under the conditions of the preceding problem? *Ans.* Letting $\varphi - \varphi_0 = x/a$, where x is the distance from the source along an arc of a circle, it will be found that the various terms are proportional to $\cos \chi x$, where χ varies in increasingly smaller steps. The expressions will tend to become integrals of certain functions with respect to χ. These integrals can be obtained directly, using the ideas expounded in Chapter 16.

3. Calculate the characteristic values and functions corresponding to the following boundary conditions: V vanishes at $\rho = a$, $\partial V/\partial \rho$ vanishes at $\rho = b$.

Ans. $n = (2m + 1)i\pi/2 \log (b/a), m = 0, 1, 2, \cdots$
$\sinh [n \log (\rho/a)]$ or $\cosh [n \log (\rho/b)]$.

4. Assume a perfectly conducting radial partition at $\varphi = \pi$. Show that the characteristic functions corresponding to the boundary condition $V(\rho,\pi) = 0$ are $\sin m\varphi$ and $\cos (m + \frac{1}{2})\varphi$ where m is an integer.

5. Show that if the radial partition of Problem 4 is at $\varphi = 0$, the characteristic functions are $\sin (m\varphi/2)$, where m is an integer.

8. *On the method of characteristic functions*

The solution of boundary value problems given in the preceding section is typical of the method of characteristic functions. The characteristic functions are solutions of a homogeneous equation; but they enable us to solve the nonhomogeneous equation as well. This is accomplished by subdividing the source function into infinitely thin layers coinciding with coordinate surfaces. For each layer we have a homogeneous equation everywhere except inside the layer itself. Solutions are written for the regions on the opposite sides of the layer and then joined at the layer. The final solution is obtained by integration. There are several options for subdividing the source function into layers. In the problem considered in the preceding section, the source function may be subdivided either in

cylindrical layers coaxial with the cylinders or in radial plane layers emerging from the common axis of the cylinders, Figure 13.2. Different sets of characteristic functions correspond to these two optional subdivisions. In the three dimensional case we should get a third set of characteristic functions corresponding to the third family of coordinate surfaces. In wave problems there is a fourth set of characteristic functions corresponding to the time variable; physically this set of characteristic functions represents the natural or free oscillations.

In some problems, the sets of characteristic values and functions are discrete; and the corresponding solutions are given by infinite series. In other problems, these sets are continuous and the series are replaced by integrals. Such problems are considered in the Chapter on Linear Analysis.

In many problems, discrete sets of characteristic functions are orthogonal; but this is not always the case.

Characteristic functions are determined by the equation and by *selected* boundary conditions. In the case of Laplace's equation in the region bounded by perfectly conducting coaxial cylinders and radial planes as in Figure 13.3, we may seek either those solutions which vanish on the radial half-planes or those which vanish on the cylinders. There are no solutions, analytic in the entire region, which vanish simultaneously on all four surfaces (barring the trivial solution $V = 0$). The solutions which vanish on all four boundaries must somewhere fail to be analytic; they may be discontinuous or have some discontinuous derivatives; or they may be infinite at some points. These mathematical singularities correspond to physical "sources" of the dependent variable.

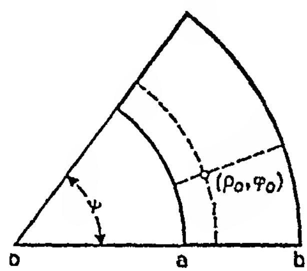

FIG. 13.3. A cross section of a cylinder whose boundaries are portions of two coaxial circular cylinders and of two half-planes emerging from the common axis. Dotted lines indicate subdivisions into subregions in which the boundary conditions of the given problem may be satisfied.

In the case of the wave equation *it is possible* to find solutions which vanish on all four boundaries of Figure 13.3, for the third independent variable t will bring in another parameter, the frequency. Such solutions will exist for certain characteristic values of this parameter, the so-called natural frequencies. The elementary "source" for expansion in this set of characteristic functions is a *pulse*. This source is analogous to the electrically charged elementary strip of the preceding section; the strip is a kind of "spatial pulse."

The method of characteristic functions owes much to the separation of

variables, which makes it easy to satisfy appropriate types of boundary conditions. This separation is not always possible. There are very few coordinate systems in which the variables of the wave equation may be separated; there are more such coordinate systems in the case of Laplace's equation; but there are many systems in which the separation of variables is impossible, and the method of characteristic functions ceases to be serviceable.

Problems

1. Show that if V is to vanish for $\varphi = 0,\psi$ the parameter n in (75) should assume one of the values given by $m\pi/\psi$, where m is an integer. Show that the corresponding characteristic functions are $\sin(m\pi\varphi/\psi)$.

2. Suppose that a charged filament passes through (ρ_0,φ_0) inside a conducting cylinder whose cross section is shown in Figure 13.3. Obtain the potential in the forms analogous to (82) and (89).

3. Suppose that a conducting rod has the boundaries shown in Figure 13.3 and the boundary $\varphi = 0$ is a perfect conductor. Let the source of current be a filament at (ρ_0,φ_0). Obtain the potential.

Note: At a perfectly conducting surface the tangential derivative of the potential vanishes; at other boundaries the normal derivative of the potential, which is proportional to the normal current density, vanishes. If the source of current is a surface, the potential and its tangential derivative are continuous but the sum of the normal derivatives in directions pointing *toward* the source is the outgoing current per unit area divided by the conductivity.

9. *Two-dimensional wave equation*

The simplest example of a two-dimensional wave equation is the equation satisfied by the displacement of an oscillating membrane. In cartesian coordinates the equation is (6); in this form it is particularly suited to rectangular membranes. If the boundaries of the membrane are radii and concentric circles, as in Figure 13.3, polar coordinates are more suitable. In these coordinates the wave equation becomes

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \varphi^2} = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}. \tag{95}$$

Assuming characteristic functions in the product form

$$u(\rho,\varphi,t) = R(\rho)\Phi(\varphi)T(t), \tag{96}$$

substituting in (95) and dividing by u, we have

$$\frac{1}{R}\frac{d^2R}{d\rho^2} + \frac{1}{\rho R}\frac{dR}{d\rho} + \frac{1}{\rho^2\Phi}\frac{d^2\Phi}{d\varphi^2} = \frac{1}{v^2T}\frac{d^2T}{dt^2}. \tag{97}$$

The left side is independent of t and the right side is a function of t only;

therefore

$$\frac{d^2T}{dt^2} = kv^2T, \tag{98}$$

where k is some constant. Substituting this constant in (97) and multiplying by ρ^2, we have

$$\frac{\rho^2}{R}\frac{d^2R}{d\rho^2} + \frac{\rho}{R}\frac{dR}{d\rho} + \frac{1}{\Phi}\frac{d^2\Phi}{d\varphi^2} = k\rho^2. \tag{99}$$

Hence,

$$\frac{d^2\Phi}{d\varphi^2} = -n^2\Phi, \tag{100}$$

$$\rho^2\frac{d^2R}{d\rho^2} + \rho\frac{dR}{d\rho} - (n^2 + k\rho^2)R = 0, \tag{101}$$

where n is another separation constant. We might have written simply "p" instead of "$-n^2$"; we are just anticipating that eventually we shall find that p should be a negative real number and that we shall want to use square roots of the parameter.

If the edges of the membrane are fixed, u must vanish at $\varphi = 0,\psi$; therefore

$$\Phi_m(\varphi) = A_m \sin(m\pi\varphi/\psi), \qquad n = m\pi/\psi. \tag{102}$$

Solutions of (101) are Bessel functions (Chapter 20), and

$$R_m(\rho) = B_mJ_n(i\sqrt{k}\rho) + C_mN_n(i\sqrt{k}\rho). \tag{103}$$

If u is to vanish at $\rho = a$ and $\rho = b$, the separation constant k must be a root of

$$-\frac{C_m}{B_m} = \frac{J_n(i\sqrt{k}a)}{N_n(i\sqrt{k}a)} = \frac{J_n(i\sqrt{k}b)}{N_n(i\sqrt{k}b)}. \tag{104}$$

The properties of Bessel functions are such that if k is to satisfy this equation, it must be negative. In setting the right side in (97) equal to k, we pretended ignorance; we might have anticipated that $T(t)$ would be a sinusoidal function of frequency ω radians per second, with ω essentially real; and this would have led to the value $k = -\omega^2/v^2$; in which case (104) would have assumed the following form

$$\frac{J_n(\omega a/v)}{N_n(\omega a/v)} = \frac{J_n(\omega b/v)}{N_n(\omega b/v)}. \tag{105}$$

This illustrates an assertion made in the preceding section: for certain

frequencies (the roots of the above equation) it is possible to find characteristic functions which vanish along the entire boundary.

Suppose now that the membrane is driven with a frequency which does not satisfy (105). The driving force introduces an extra term in the wave equation (95) and makes it nonhomogeneous. Just as in the case of forced vibrations of a string, we can express the motion of the membrane in terms of the free vibrations; but instead of repeating the routine, we might just as well consider the more general case, independently of a particular coordinate system (see Section 11).

Two other types of expression for forced oscillations are analogous to the expansions in Section 7. The only difference is that R is now a linear combination of Bessel functions J_n, N_n instead of power functions ρ^n, ρ^{-n}. As the frequency of forced oscillations approaches zero, J_n approaches ρ^n (except for a constant factor) and N_n approaches ρ^{-n}.

10. *Three-dimensional wave equation*

If there are four independent variables, three coordinates of a point, and time, there will be three separation constants and three sets of characteristic constants. Consider the wave equation in spherical coordinates

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\varphi^2}=\frac{1}{v^2}\frac{\partial^2 V}{\partial t^2}. \tag{106}$$

Assuming

$$V(r,\theta,\varphi;t) = R(r)\Theta(\theta)\Phi(\varphi)T(t) \tag{107}$$

and separating the variables, we have

$$\frac{d^2T}{dt^2} = -\omega^2 T, \qquad \frac{d^2\Phi}{d\varphi^2} = -\mu^2\Phi, \tag{108}$$

$$\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+[\nu(\nu+1)\sin^2\theta-\mu^2]\Theta=0, \tag{109}$$

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)=[\nu(\nu+1)-\beta^2r^2]R, \qquad \beta=\omega/v. \tag{110}$$

Equation (109) is the associated Legendre equation; (110) can be reduced to the Bessel equation by setting R equal to $w/\sqrt{r}$, where w is the new dependent variable, ($R\sqrt{r}$ is a Bessel function). An even more suitable transformation is $R = \hat{R}/r$; this transformation removes the first derivative in (110) and leads to

$$\frac{d^2\hat{R}}{dr^2} = -\beta^2\hat{R}+\frac{\nu(\nu+1)}{r^2}\hat{R}. \tag{111}$$

As r approaches infinity, $\hat{R}$ becomes nearly sinusoidal.

The reason for writing one of the characteristic constants as $\nu(\nu+1)$ is largely historical. If μ and ν are both integers, one solution of the associated Legendre equation can be expressed in closed form in terms of circular functions. In particular, when $\mu = 0$ the solution is a polynomial in $\cos\theta$; this solution is denoted by $P_n(\cos\theta)$, $(\nu = n)$. In general, however, μ and ν need not be integers, or even real. Modern applications require general solutions of the above equations. Habitually Bessel and Legendre functions are considered as functions of one independent variable, the variable appearing as the independent variable in the differential equations; it is best, however, to regard Bessel and Legendre functions as functions of two variables, including the characteristic "constants."

Functions of the form (107) are particularly suitable for regions bounded by pairs of coordinate surfaces,

$$r = r_1 \text{ and } r = r_2;\quad \theta = \theta_1 \text{ and } \theta = \theta_2;\quad \varphi = \varphi_1 \text{ and } \varphi = \varphi_2. \tag{112}$$

It becomes easy to make either V or its normal derivative vanish at any particular boundary and thus satisfy some of the most frequently occurring boundary conditions. If V has to vanish at the half-planes $\varphi = \varphi_1, \varphi_2$, μ is immediately restricted and becomes equal to $m\pi/(\varphi_2 - \varphi_1)$, where m is an integer. Similarly, if V is to vanish at the conical surfaces $\theta = \theta_1, \theta_2$, μ and ν have to satisfy a certain equation involving Legendre's functions. Since μ has already been determined, this equation restricts ν to a certain set of values. Finally if V is to vanish at the spherical boundaries $r = r_1, r_2$, β and ν have to satisfy a certain equation involving solutions of (111). Since ν has already been restricted, β and therefore ω becomes restricted to a certain set of values.

If ω is given and happens to be different from one of the characteristic values, the homogeneous equation has no solution; instead the nonhomogeneous equation, with the "source function," has a solution which can be expressed as a series of functions of the form (107).

There are other alternatives. We can use the fact that ω has been fixed by the conditions of the problem and relax the boundary conditions *at one pair* of surfaces given by (112). These temporarily ignored conditions are later satisfied by a *discontinuous* solution of the form (107) from which the solution of the nonhomogeneous equation is obtained by summation or integration.

In the following sections these processes are carried out independently of any particular coordinate system. The problem is then reduced essentially to the calculation of characteristic functions, and ultimately to the determination of solutions of second-order linear homogeneous ordinary differential equations, of which the Bessel and Legendre equations are only special cases.

Problems

1. Show that the wave equation is satisfied by $A \exp(pt - \gamma_x x - \gamma_y y - \gamma_z z)$, where the *oscillation constant* p and *propagation constants* $\gamma_x, \gamma_y, \gamma_z$ in the directions of the coordinate axes may be complex provided they satisfy the following equation

$$\gamma_x^2 + \gamma_y^2 + \gamma_z^2 = p^2/v^2.$$

2. Suppose that physical requirements are such that the wave function should vanish at the boundaries of a parallelepiped whose dimensions are a, b, c. Find the natural frequencies.

Ans. $\omega_{\ell,m,n} = 2\pi f_{\ell,m,n} = v\pi\sqrt{(\ell/a)^2 + (m/b)^2 + (n/c)^2}$,

where $\ell, m, n = 1, 2, \cdots$.

3. Separate the variables in the wave equation in cylindrical coordinates, in elliptic cylinder coordinates, and in spheroidal coordinates.

11. *Cavity resonators — scalar oscillations*

Let us now write the wave equation in a form independent of any particular coordinate system

$$\Delta V = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2}. \tag{113}$$

Setting

$$V(x,y,z;t) = \psi(x,y,z)T(t), \tag{114}$$

we obtain

$$\Delta\psi = -\beta^2\psi, \qquad \beta = \omega/v, \tag{115}$$

$$\frac{d^2T}{dt^2} = -\omega^2 T, \qquad T = A\cos(\omega t + \vartheta). \tag{116}$$

Applying Green's formula (7–61) to $U = V = \psi$, we get

$$-\beta^2 \iiint \psi^2\, d\tau = \iint \psi\frac{\partial\psi}{\partial n}\, dS - \iiint (\operatorname{grad}\psi)^2\, d\tau. \tag{117}$$

Frequently either ψ or $\partial\psi/\partial n$ is required to vanish at the boundary; then

$$\beta^2 = \frac{\omega^2}{v^2} = \frac{\iiint (\operatorname{grad}\psi)^2\, d\tau}{\iiint \psi^2\, d\tau}. \tag{118}$$

This is a three-dimensional analogue of (11–110).

If v is a function of position, $1/v$ must be retained under the integral sign

$$\omega^2 = \frac{\iiint (\text{grad}\,\psi)^2\, d\tau}{\iiint (\psi/v)^2\, d\tau}. \tag{119}$$

To generalize still further, we replace (115) by

$$\text{div}\,(P\,\text{grad}\,\psi) = -\omega^2 Q\psi, \tag{120}$$

where P and Q are functions of position; then

$$\omega^2 = \frac{\iiint P(\text{grad}\,\psi)^2\, d\tau}{\iiint Q\psi^2\, d\tau}. \tag{121}$$

It is just as easy to deal with the general wave equation (120) as with its special case (115). With this in mind we shall generalize Green's formula (7–61). First we note that

$$\text{div}\,(\psi_1 P\,\text{grad}\,\psi_2) = \psi_1\,\text{div}\,(P\,\text{grad}\,\psi_2) + P(\text{grad}\,\psi_1)\cdot(\text{grad}\,\psi_2), \tag{122}$$

for any three differentiable scalar functions (P has to be differentiable only once but ψ_1 and ψ_2 should be twice differentiable). Integrating over the volume bounded by a closed surface S, we have

$$\iiint \psi_1\,\text{div}\,(P\,\text{grad}\,\psi_2)\, d\tau + \iiint P\,(\text{grad}\,\psi_1)\cdot(\text{grad}\,\psi_2)\, d\tau = \iint P\psi_1 \frac{\partial\psi_2}{\partial n}\, dS, \tag{123}$$

where $\partial/\partial n$ indicates differentiation along the outward normal to S. Interchanging ψ_1 and ψ_2 and subtracting, we obtain

$$\iiint [\psi_1\,\text{div}\,(P\,\text{grad}\,\psi_2) - \psi_2\,\text{div}\,(P\,\text{grad}\,\psi_1)]\, d\tau = \iint P\left(\psi_1\frac{\partial\psi_2}{\partial n} - \psi_2\frac{\partial\psi_1}{\partial n}\right) dS. \tag{124}$$

In some physical problems the properties of the oscillating medium undergo an abrupt change at a given surface; it can be shown that such discontinuities do not affect the validity of the above equation.

Characteristic functions belonging to different characteristic values of

ω are orthogonal, with Q as the weight function,

$$\iiint Q\psi_1\psi_2 \, d\tau = 0 \text{ if } \omega_1 \neq \omega_2. \tag{125}$$

The proof follows immediately if we substitute from (120) into (124).

There may exist several linearly independent characteristic functions corresponding to the same characteristic value of ω. The foregoing proof breaks down in this case, and indeed the characteristic functions as obtained in the natural course of solution may be nonorthogonal. It is possible, however, to replace them by an orthogonal set. Suppose that

$$\iiint Q\psi_1\psi_2 \, d\tau \neq 0. \tag{126}$$

Consider a linear combination

$$\psi = A\psi_1 + \psi_2. \tag{127}$$

It is possible to determine A such that ψ and ψ_1 are orthogonal; thus

$$\iiint Q\psi_1\psi \, d\tau = A\iiint Q\psi_1^2 \, d\tau + \iiint Q\psi_1\psi_2 \, d\tau = 0,$$

$$A = -\frac{\iiint Q\psi_1\psi_2 \, d\tau}{\iiint Q\psi_1^2 \, d\tau}. \tag{128}$$

Equations (121) and (128) break down if the denominators vanish. In physical applications this does not happen because P and Q are positive throughout the volume.

Step by step, any nonorthogonal set can be converted into an orthogonal set; thus if ψ_1 and ψ_2 are orthogonal while ψ_3 is not, we write $\psi = A_1\psi_1 + A_2\psi_2 + \psi_3$ and determine A_1 and A_2 to make ψ orthogonal to ψ_1 and ψ_2.

The set of characteristic values is a discrete set. To prove it, let us assume the opposite, that all values of ω in a given interval are characteristic values. Let $\psi(x,y,z;\omega)$ be a characteristic function depending on the parameter ω; then

$$\psi(x,y,z;\omega + \delta\omega) = \psi(x,y,z;\omega) + \frac{\partial\psi}{\partial\omega}\,\delta\omega + \cdots. \tag{129}$$

The orthogonality property leads to

$$\iiint Q\psi^2 \, d\tau + \delta\omega\iiint Q\psi\frac{\partial\psi}{\partial\omega} \, d\tau + \cdots = 0. \tag{130}$$

We have assumed that $\delta\omega$ can be as small as we wish; hence the first integral vanishes, and that *is impossible as long as Q is of the same sign throughout the volume.*

In the foregoing analysis there is another obvious assumption: *the volume bounded by S is finite.* This assumption makes the integrals finite (assuming of course that Q is finite). If the integrals are not finite, our proofs break down. For infinite regions characteristic values *may* form a continuous set. For example, the natural frequencies of a box become closer as the size of the box increases which suggests that for an infinitely large box every frequency is a " natural frequency "; this is indeed the case.

Next let us consider the nonhomogeneous equation

$$\operatorname{div}(P \operatorname{grad} \psi) = -\omega^2 Q\psi + F, \tag{131}$$

where F is a given function of position. Suppose

$$F = \sum A_n Q\psi_n, \qquad \psi = \sum B_n \psi_n, \tag{132}$$

where the ψ_n's form a set of orthogonal characteristic functions of the associated homogeneous equation. Multiplying the first equation by ψ_m and integrating over the volume enclosed by S, we find

$$A_m = \frac{\iiint F\psi_m \, d\tau}{\iiint Q\psi_m^2 \, d\tau}. \tag{133}$$

Substituting ψ and F in (131), we obtain

$$\begin{aligned} -\sum B_n \omega_n^2 Q\psi_n &= -\omega^2 \sum B_n Q\psi_n + \sum A_n Q\psi_n, \\ B_n &= A_n/(\omega^2 - \omega_n^2). \end{aligned} \tag{134}$$

Thus we have the following particular solution of the nonhomogeneous equation

$$\psi = \sum \frac{\psi_n \iiint F\psi_n \, d\tau}{(\omega^2 - \omega_n^2)\iiint Q\psi_n^2 \, d\tau}. \tag{135}$$

This " response " is of the same frequency as the source function, and is called the *steady state* response. To this we may add a general solution of the homogeneous equation, that is, the sum of characteristic functions with arbitrary amplitudes; this is the *complementary solution* or the *transient response.* In physical systems it gradually dies out on account of dissipation.

Just as in the one-dimensional case, (135) represents the true answer only if the set of characteristic functions is complete. In problems involving cartesian, cylindrical and spherical coordinates complete sets are obtained in the course of solution; but proofs of completeness are not simple.

Characteristic functions are said to be normalized if

$$\iiint Q\psi_n^2\,d\tau = 1. \tag{136}$$

The obvious advantage is the simplification of the form of such equations as (133) and (135).

If F is zero everywhere except in an infinitesimal region surrounding a point (x_0,y_0,z_0) where it is infinitely large in such a way that $\iiint F\,d\tau = 1$, then

$$\psi = G(x,y,z;x_0,y_0,z_0) = \sum \frac{\psi_n(x,y,z)\psi_n(x_0,y_0,z_0)}{\omega^2 - \omega_n^2}. \tag{137}$$

This is the Green's function for our boundary value problem.

So far it has been assumed that the source function is of frequency ω. In the Chapter on Linear Analysis it is shown that any function which is likely to be of interest in physical applications may be represented either as a series or an integral of the sinusoidal source function; thus, at least in theory, we have a very general solution.

12. *Cavity resonators — vector oscillations*

Vector oscillations occur in electromagnetic problems in which the equations connect two vectors, the electric vector E and magnetic vector H,

$$\operatorname{curl} E = -i\omega\mu H, \qquad \operatorname{curl} H = i\omega\epsilon E. \tag{138}$$

These equations are obtained from Maxwell's equations by assuming that the instantaneous values of the electric and magnetic vectors are real parts of $E\exp(i\omega t)$ and $H\exp(i\omega t)$.

Normally μ and ϵ are assumed to be constant; but in the following analysis we shall regard them as functions of position. Either E or H can be eliminated from (138):

$$\operatorname{curl}(\mu^{-1}\operatorname{curl} E) = \omega^2\epsilon E, \tag{139}$$

$$\operatorname{curl}(\epsilon^{-1}\operatorname{curl} H) = \omega^2\mu H. \tag{140}$$

In a region bounded by a perfectly conducting surface the tangential component of E vanishes at the surface. What follows could be applied

equally well to the hypothetical cases in which the tangential component of H vanishes either over the entire surface or a part of it.

In (138) E and H may be complex, and in order to insure that the denominators in such equations as (121) do not vanish we shall have to introduce conjugate complex vectors E^* and H^*. As in the case of the scalar wave equation, we shall assume that μ and ϵ are positive throughout the resonator.

If the first equation in (138) is multiplied scalarly by H^*, and the conjugate of the second by E, then the difference is

$$\begin{aligned} i\omega(\epsilon E \cdot E^* - \mu H \cdot H^*) &= H^* \cdot \text{curl } E - E \cdot \text{curl } H^* \\ &= \text{div } (E \times H^*). \end{aligned} \tag{141}$$

Integrating this over the volume bounded by S, we find that the right-hand side vanishes if the tangential component of either E or H vanishes at the boundary. Thus the volume integral of the divergence equals the surface integral of the normal component of $E \times H^*$; this component vanishes if $E \times H^*$ is tangential to S; since this vector is normal to E and H, it will automatically be tangential to S if either E or H is normal to S. Thus we obtain

$$\iiint \epsilon E \cdot E^* \, d\tau = \iiint \mu H \cdot H^* \, d\tau. \tag{142}$$

Physically this equation means the equality of the average electric and magnetic energies in the field.

With the aid of (138) either E or H can be eliminated; then

$$\begin{aligned} \omega^2 &= \frac{\displaystyle\iiint \epsilon^{-1} \, (\text{curl } H) \cdot (\text{curl } H^*) \, d\tau}{\displaystyle\iiint \mu H \cdot H^* \, d\tau} \\ &= \frac{\displaystyle\iiint \mu^{-1} \, (\text{curl } E) \cdot (\text{curl } E^*) \, d\tau}{\displaystyle\iiint \epsilon E \cdot E^* \, d\tau}. \end{aligned} \tag{143}$$

Next we shall prove that the characteristic functions corresponding to different frequencies are orthogonal

$$\iiint \epsilon E_1 \cdot E_2^* \, d\tau = \iiint \mu H_1 \cdot H_2^* \, d\tau = 0. \tag{144}$$

Physically these equations mean that there is no coupling between different

oscillation modes, that the mutual energies vanish. The proof is not difficult once we realize that it must depend on the boundary conditions, that is, on the vanishing of the vector product of two fields. Thus by (7–56) and (138) we have

$$\begin{aligned} \operatorname{div}(E_1 \times H_2^*) &= H_2^* \cdot \operatorname{curl} E_1 - E_1 \cdot \operatorname{curl} H_2^* \\ &= -i\omega_1 \mu H_1 \cdot H_2^* + i\omega_2 \epsilon E_1 \cdot E_2^*. \end{aligned} \tag{145}$$

Interchanging the subscripts and taking the conjugate, we have

$$\operatorname{div}(E_2^* \times H_1) = i\omega_2 \mu H_1 \cdot H_2^* - i\omega_1 \epsilon E_1 \cdot E_2^*. \tag{146}$$

Multiplying (145) by ω_2, (146) by ω_1, and adding

$$i(\omega_2^2 - \omega_1^2)\epsilon E_1 \cdot E_2^* = \omega_2 \operatorname{div}(E_1 \times H_2^*) + \omega_1 \operatorname{div}(E_2^* \times H_1). \tag{147}$$

Integrating over the volume bounded by S, we find that E_1 and E_2^* are orthogonal as long as either one or the other tangential component vanishes over S.

Characteristic functions belonging to the same value of ω are not necessarily orthogonal; but they may be replaced by an equal number of orthogonal functions. The proof is essentially the same as in the case of scalar functions. Similarly it can be shown that the characteristic values form a discrete set.

The physical source function of an electromagnetic field is either the conduction or the convection current. In the presence of a source, Maxwell's equations become

$$\operatorname{curl} E = -i\omega\mu H, \qquad \operatorname{curl} H = i\omega\epsilon E + J, \tag{148}$$

where J is the density of electric current. Instead of (139) we now have

$$\operatorname{curl}(\mu^{-1} \operatorname{curl} E) = \omega^2 \epsilon E - i\omega J. \tag{149}$$

Let

$$J = \sum A_n \epsilon E_n, \qquad E = \sum B_n E_n, \tag{150}$$

where the summation is extended over the complete normalized set of characteristic functions of the associated homogeneous equations; then

$$A_n = \iiint J \cdot E_n^* \, d\tau, \qquad B_n = i\omega A_n/(\omega^2 - \omega_n^2). \tag{151}$$

Hence,

$$E = \sum \frac{i\omega}{\omega^2 - \omega_n^2} E_n \iiint J \cdot E_n^* \, d\tau. \tag{152}$$

For an electric current element of unit moment at (x_0,y_0,z_0) (152)

becomes

$$E(x,y,z;x_0,y_0,z_0) = \sum \frac{i\omega}{\omega^2 - \omega_n^2} E_n^*(x_0,y_0,z_0)E_n(x,y,z). \tag{153}$$

The corresponding expression for the H-vector is

$$H(x,y,z;x_0,y_0,z_0) = \sum \frac{i\omega_n}{\omega^2 - \omega_n^2} E_n^*(x_0,y_0,z_0)H_n(x,y,z). \tag{154}$$

A similar solution may be obtained for

$$\operatorname{curl} E = -i\omega\mu H - C, \qquad \operatorname{curl} H = i\omega\epsilon E, \tag{155}$$

where C is the density of magnetic current. This is a nonphysical case, but of considerable importance in theory because it simplifies the treatment of circulating current and of double layers of electric current. Thus, for a magnetic current element of unit moment, we obtain

$$\begin{aligned} H(x,y,z;x_0,y_0,z_0) &= \sum \frac{i\omega}{\omega^2 - \omega_n^2} H_n^*(x_0,y_0,z_0)H_n(x,y,z), \\ E(x,y,z;x_0,y_0,z_0) &= \sum \frac{i\omega_n}{\omega^2 - \omega_n^2} H_n^*(x_0,y_0,z_0)E_n(x,y,z). \end{aligned} \tag{156}$$

The determination of complete sets of characteristic functions is relatively easy if the boundaries are the coordinate surfaces in cartesian, cylindrical and spherical frames of reference, because the variables in the equations are separable. A few other cases are manageable for the same reason; but in most cases the spatial variables in the wave equation are not separable. The formulas in this section hold, but the calculation of the characteristic functions becomes very difficult.

Problems

1. Starting with Maxwell's equations for the *instantaneous* values of E and H,

$$\operatorname{curl} E = -\mu\frac{\partial H}{\partial t}, \qquad \operatorname{curl} H = \epsilon\frac{\partial E}{\partial t},$$

show that

$$\tfrac{1}{2}\iiint \epsilon E\cdot E\,d\tau + \tfrac{1}{2}\iiint \mu H\cdot H\,d\tau = \text{constant};$$

that is, the total energy content of a resonator is constant.

2. Determine the complete set of characteristic functions for an electric cavity resonator in the shape of a parallelepiped. The boundary condition is that the tangential component of E vanishes at the boundary.

Ans.

$$E_z = A\chi^2 \sin\frac{\ell\pi x}{a}\sin\frac{m\pi y}{b}\cos\frac{n\pi z}{c}\exp(i\omega t),$$

$$H_z = B\chi^2 \cos\frac{\ell\pi x}{a}\cos\frac{m\pi y}{b}\sin\frac{n\pi z}{c}\exp(i\omega t),$$

$$E_x = \left(-\frac{\ell n\pi^2}{ac}A + i\omega\mu\frac{m\pi}{b}B\right)\cos\frac{\ell\pi x}{a}\sin\frac{m\pi y}{b}\sin\frac{n\pi z}{c}\exp(i\omega t),$$

$$E_y = \left(-\frac{mn\pi^2}{bc}A - i\omega\mu\frac{\ell\pi}{a}B\right)\sin\frac{\ell\pi x}{a}\cos\frac{m\pi y}{b}\sin\frac{n\pi z}{c}\exp(i\omega t),$$

$$H_x = \left(i\omega\epsilon\frac{m\pi}{b}A - \frac{\ell n\pi^2}{ac}B\right)\sin\frac{\ell\pi x}{a}\cos\frac{m\pi y}{b}\cos\frac{n\pi z}{c}\exp(i\omega t),$$

$$Hy = \left(-i\omega\epsilon\frac{\ell\pi}{a}A - \frac{mn\pi^2}{bc}B\right)\cos\frac{\ell\pi x}{a}\sin\frac{m\pi y}{b}\cos\frac{n\pi z}{c}\exp(i\omega t),$$

where

$$\chi^2 \equiv \chi^2_{\ell,m} = \frac{\ell^2\pi^2}{a^2} + \frac{m^2\pi^2}{b^2}, \qquad \chi^2 + \frac{n^2\pi^2}{c^2} = \beta^2 = \omega^2\mu\epsilon, \qquad \omega \equiv \omega_{\ell,m,n},$$

and the constants A and B should also be written $A_{\ell,m,n}, B_{\ell,m,n}$.

13. *Wave guides*

In the preceding analysis of cavity resonators it has been assumed that their dimensions are finite. If one dimension becomes infinite, we have a *wave guide;* and, in general, the analysis has to be modified. For instance, we are no longer justified in assuming that the integrals in (143) are finite. In fact, the proof of this formula breaks down; for it is based on the vanishing of a certain integral extended over the boundary of the resonator. Of course, we can isolate a part of a wave guide from the rest and apply Green's theorem; but we shall find no reason for supposing that the integral over the boundary vanishes. Thus we have to reconsider the entire analysis.

Suppose that we are interested in the solutions of the following scalar wave equation

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = -\omega^2\mu\epsilon\psi, \tag{157}$$

suitable to the interior of a cylindrical shell whose generators are parallel to the z-axis. We shall assume that μ and ϵ are independent of the coordinates. This equation is satisfied by the cartesian components of the E and H vectors in any region free from conduction and convection currents.

In particular if the boundary is a perfect conductor, and if $\psi = E_z$, then ψ vanishes on the boundary; if $\psi = H_z$, then the normal derivative $\partial\psi/\partial n$ vanishes on the boundary.

Substituting

$$\psi = U(x,y)Z(z) \tag{158}$$

in (157) and separating the longitudinal coordinate from the transverse ones, we obtain

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = -\chi^2 U, \tag{159}$$

$$\frac{d^2 Z}{dz^2} = (\chi^2 - \omega^2\mu\epsilon)Z, \tag{160}$$

where the separation constant χ is so far arbitrary. If either ψ or $\partial\psi/\partial n$ is to vanish at the boundary for all values of z, U or $\partial U/\partial n$ must vanish there. Our problem becomes essentially the same as that solved in Section 11, except that we now have two dimensions instead of three. The volume and surface integrals will thus be replaced by surface and line integrals. The parameter χ will be restricted to certain characteristic values. There being no boundary conditions on Z, the frequency ω is unrestricted. In an infinite region the set of characteristic values of ω is a continuous set; all values are permissible. The series expansions of Section 11 are not applicable to ψ; but similar expansions apply to U.

The general expression for Z is

$$Z(z) = C\exp\left(z\sqrt{\chi^2 - \omega^2\mu\epsilon}\right) + D\exp\left(-z\sqrt{\chi^2 - \omega^2\mu\epsilon}\right). \tag{161}$$

To the right of all sources only the second term is present; for if the coefficient of z is real, the first term would increase with the distance from the source — a physically impossible situation. If the coefficient of z is imaginary, the phase delay must increase with the distance from the source, or else a wave would exist before its cause. Similarly, only the first term is present in the region to the left of all sources.

The analogue of (118) is

$$\chi^2 = \frac{\displaystyle\iint (\operatorname{grad} U)^2\, dS}{\displaystyle\iint U^2\, dS}, \tag{162}$$

where the integration is extended over the cross section of the guide. The orthogonality property becomes

$$\iint U_1 U_2\, dS = 0 \text{ if } \chi_1 \neq \chi_2. \tag{163}$$

Once the complete set of characteristic functions has been determined, it is easy to express the wave function in terms of its sources. If, for instance, the sources are in the plane $z = 0$, we have

$$\begin{aligned} \psi &= \sum A_n U_n \exp\left(-z\sqrt{\chi_n^2 - \omega^2\mu\epsilon}\right), \qquad z \geq 0; \\ &= \sum B_n U_n \exp\left(z\sqrt{\chi_n^2 - \omega^2\mu\epsilon}\right), \qquad z \leq 0. \end{aligned} \tag{164}$$

Suppose that ψ is continuous and $\partial\psi/\partial z$ discontinuous, the discontinuity being given by $F(x,y)$. Then $A_n = B_n$ and

$$\left.\frac{\partial\psi}{\partial z}\right|_{z=+0} - \left.\frac{\partial\psi}{\partial z}\right|_{z=-0} = F(x,y) = -2\sum A_n\sqrt{\chi_n^2 - \omega^2\mu\epsilon}\,U_n. \tag{165}$$

Since the U's are orthogonal, we find

$$A_n = -\frac{\int\int FU_n\,dS}{2\sqrt{\chi_n^2 - \omega^2\mu\epsilon}\int\int U_n^2\,dS}. \tag{166}$$

If the discontinuity is restricted to an infinitely small area, if $\int\int F\,dS = 1$, and if the characteristic functions are normalized, then

$$A_n = -\frac{U_n(x_0,y_0)}{2\sqrt{\chi_n^2 - \omega^2\mu\epsilon}}. \tag{167}$$

Some differences should be noted in the form of the present answer as compared to the forms in the previous sections. The denominators do not contain factors $\chi^2 - \chi_n^2$ which would be analogous to $\omega^2 - \omega_n^2$. Previous sections deal with forced oscillations so that ω is prescribed. This section deals with "free" or "natural waves." Of course, the wave is started somewhere in the guide; but the sources are localized and everywhere else the wave motion is governed solely by the boundary constraints. "Forced waves" analogous to forced oscillations will occur if we insert in the wave equation a source function of the form $P\exp(-i\chi z)$, in which the phase constant χ is given (see problems at the end of this section).

If we assume two conducting partitions at $z = 0$, l, we shall have a cavity resonator. The idea of regarding the cavity resonator as a section of a wave guide leads to another form of solution in which the resonator response is given as the sum of progressive waves emitted from the source and of waves reflected from the ends of the guide. The essential difference may be brought out by comparing the solutions for a simple one-dimensional case.

The voltage V and current I in a uniform transmission line obey the following equations

$$\frac{dV}{dx} = -i\omega LI + E(x), \qquad \frac{dI}{dx} = -i\omega CV, \tag{168}$$

where L, C are constant and $E(x)$ is the impressed series voltage per unit length. An important special case is the one in which $E(x)$ is concentrated at $x = \xi$ in such a way that its integral over the interval $(\xi - 0, \xi + 0)$ is unity. This makes V discontinuous

$$V(\xi + 0) - V(\xi - 0) = 1. \tag{169}$$

The current is continuous.

If the line is shorted at $x = 0$ and $x = \ell$, then

$$V(0) = V(\ell) = 0, \tag{170}$$

and we have a one-dimensional resonator. We shall *now solve this problem* in two ways. First we shall take advantage of the fact that in the present problem $E(x)$ vanishes everywhere except at $x = \xi$; hence $V(x)$ is a solution of homogeneous equations everywhere except at $x = \xi$. Thus

$$\begin{aligned} V(x) &= A \sin (\omega\sqrt{LC}x), \qquad x < \xi; \\ &= B \sin [\omega\sqrt{LC}(\ell - x)], \; x > \xi. \end{aligned} \tag{171}$$

One part of the solution satisfies the required boundary condition at $x = 0$; the other at $x = \ell$. The constants A, B may be found to satisfy (169) and make the current continuous. Thus (169) becomes

$$B \sin [\omega\sqrt{LC}(\ell - \xi)] - A \sin (\omega\sqrt{LC}\xi) = 1. \tag{172}$$

The current in the two regions is

$$\begin{aligned} I = \frac{i}{\omega L}\frac{dV}{dx} &= Ai\sqrt{C/L} \cos (\omega\sqrt{LC}x), \qquad x < \xi; \\ &= -Bi\sqrt{C/L} \cos [\omega\sqrt{LC}(\ell - x)], \qquad x > \xi. \end{aligned} \tag{173}$$

Hence,

$$A \cos (\omega\sqrt{LC}\xi) = -B \cos [\omega\sqrt{LC}(\ell - \xi)]. \tag{174}$$

Solving for A and B and substituting in (171), we obtain

$$
\begin{aligned}
V(x) &= -\frac{\cos[\omega\sqrt{LC}(\ell-\xi)]\sin(\omega\sqrt{LC}x)}{\sin\beta\ell}, && x<\xi;\\
&= \frac{\cos(\omega\sqrt{LC}\xi)\sin[\omega\sqrt{LC}(\ell-x)]}{\sin\beta\ell}, && x>\xi;\\
I(x) &= -i\sqrt{C/L}\frac{\cos[\omega\sqrt{LC}(\ell-\xi)]\cos(\omega\sqrt{LC}x)}{\sin\beta\ell}, && x<\xi;\\
&= -i\sqrt{C/L}\frac{\cos(\omega\sqrt{LC}\xi)\cos[\omega\sqrt{LC}(\ell-x)]}{\sin\beta\ell}, && x>\xi.
\end{aligned}
\tag{175}
$$

Next let us solve the same problem in terms of the natural oscillations of the line section. For this purpose we eliminate V from (168)

$$\frac{d^2I}{dx^2} = -\omega^2 LCI - i\omega CE(x). \tag{176}$$

The characteristic functions are solutions of the associated homogeneous equation obtained by setting $E(x) = 0$ everywhere. The characteristic values are obtained from the condition that $V(x)$, and therefore dI/dx, must vanish at $x = 0$ and $x = \ell$. Thus the characteristic values are

$$\omega_n\sqrt{LC}\ell = n\pi, \qquad \omega_n = n\pi/\sqrt{LC}\ell, \qquad n = 0, 1, 2, \cdots, \tag{177}$$

and the corresponding characteristic functions are

$$I_n(x) = A_n \cos\beta_n x, \qquad \beta_n = \omega_n\sqrt{LC}. \tag{178}$$

Normalizing,

$$I_0(x) = 1/\sqrt{\ell LC}, \qquad I_n(x) = \sqrt{2/\ell LC}\cos\beta_n x \text{ if } n \neq 0, \tag{179}$$

and applying (137), we obtain

$$I(x) = \frac{1}{i\omega L\ell} - \frac{2i\omega}{L\ell}\sum_{n=1}^{\infty}\frac{\cos\beta_n\xi\cos\beta_n x}{\omega^2-\omega_n^2}. \tag{180}$$

In obtaining this expression from (137) we should note that, in order to compare the result with (175), we want the integral of $E(x)$ to be unity; therefore the integral of the source function in (176) should be $-i\omega C$. Also we should note that, according to (133), the normalization should be performed with the weight function Q. Comparing (120) and (176), we obtain $P = 1$ and $Q = LC$.

Equations of the form (175) are simpler than (180); but not every resonator is a section of a wave guide.

Problems

1. In equation (176) let the impressed series voltage per unit length, $E(x)$, be

equal to $E_0 \exp(-i\chi x)$ in the interval $(-\infty,\infty)$. Find the forced wave.

Ans. $I(x) = \dfrac{i\omega CE_0}{\chi^2 - \omega^2 LC} e^{-i\chi x}.$

2. In the preceding problem let the impressed voltage act only in the interval $(0,\ell)$. Find the current.

$$\begin{aligned} I(x) &= \frac{E_0[1 - e^{-i(\chi+\beta)\ell}]}{2iK(\chi+\beta)} e^{i\beta x}, \qquad x \le 0; \qquad K = \sqrt{\frac{L}{C}}; \\ &= \frac{i\omega CE_0}{\chi^2 - \beta^2} e^{-i\chi x} + \frac{E_0 e^{-i\beta x}}{2iK(\chi-\beta)} - \frac{E_0 e^{i\beta(x-\ell)-i\chi\ell}}{2iK(\chi+\beta)}, \qquad 0 \le x \le \ell; \\ &= \frac{E_0[1 - e^{-i(\chi-\beta)\ell}]}{2iK(\chi-\beta)} e^{-i\beta x}, \qquad x \ge \ell; \qquad \beta = \omega\sqrt{LC}. \end{aligned}$$

3. In the solution of the problem given by equations (164) to (167) assume that the source function is proportional to $e^{-i\chi z}\,dz$ in the interval $0 \le z \le \ell$ and is zero elsewhere. Find ψ. *Hint:* Integrate the solution given in the text.

Ans.
$$\begin{aligned} \psi &= \sum \frac{U_n(x_0,y_0)U_n(x,y)}{2k_n(\chi+k_n)} [1 - e^{-i(\chi+k_n)\ell}] e^{ik_n z}, \qquad z \le 0; \\ &= \sum \frac{U_n(x_0,y_0)U_n(x,y)}{k_n^2 - \chi^2} e^{-i\chi z} \\ &\quad + \sum \frac{U_n(x_0,y_0)U_n(x,y)}{2k_n} \left[\frac{e^{-ik_n z}}{\chi - k_n} - \frac{e^{-ik_n(\ell-z)-i\chi\ell}}{\chi+k_n}\right], \qquad 0 \le z \le \ell; \\ &= \sum \frac{U_n(x_0,y_0)U_n(x,y)}{2k_n(\chi-k_n)} [1 - e^{-i(\chi-k_n)\ell}] e^{-ik_n z}, \qquad z \ge \ell; \end{aligned}$$

where $k_n = \sqrt{\omega^2\mu\epsilon - \chi_n^2}$.

The forced wave is that given by the term proportional to $\exp(-i\chi z)$.

14. *Absorption of energy*

All physical systems are dissipative, although many only slightly. In some instances a general problem including dissipation is just as easily solved as the same problem without dissipation; but in others complications are so great that an approximate method must be used. If only a small fraction of the energy stored in the oscillating system is dissipated in each cycle, it is possible to convert the exact solutions for the corresponding nondissipative system into approximate solutions for the dissipative system. The method will be illustrated with an application to an electric cavity resonator bounded by an imperfect conductor.

In this case Maxwell's equations are

$$\operatorname{curl} E = -i\omega\mu H, \qquad \operatorname{curl} H = i\omega\epsilon E + gE + J, \tag{181}$$

where g is the conductivity. Let us assume that in the dielectric $g = 0$ and in the conductor g is very large. If g were infinite, the tangential electric intensity would vanish at the boundary; in the actual case the intensity will be small. It can be established by an independent analysis that the field in the conductor is attenuated very rapidly as we recede from the inner boundary along the normal. We shall assume that the conductor is sufficiently thick to make the field at its outer boundary very nearly equal to zero.

Applying to (181) the same operations which transform (138) into (141), we have

$$i\omega\iiint(\epsilon E\cdot E^* - \mu H\cdot H^*)\,d\tau - \iiint gE\cdot E^*\,d\tau - \iiint E\cdot J^*\,d\tau = 0, \tag{182}$$

where the volume integration is extended to the *outer* boundary of the resonator. In what follows we need not know the physical interpretation of this equation; but it will help to understand the results. One half of the first integral is the difference of the average electric and magnetic energies stored in the resonator; one half of the second integral is the average power dissipated in the conductor; and one half of the third integral is the average power contributed to the resonator by the applied forces. Our assumption is that the middle integral is very small. If we neglect it, we find an expansion of the form (152) and the natural frequencies are the roots of the characteristic equation

$$\iiint \epsilon E\cdot E^*\,d\tau = \iiint \mu H\cdot H^*\,d\tau. \tag{183}$$

When $\omega = \omega_n$, the amplitude of the response is infinite if $J \neq 0$. Equation (182) shows, however, that in the dissipative case the amplitude is not infinite but is determined by the power supplied to the resonator. Thus we write

$$\begin{aligned} E &= A_n i\omega_n E_n \iiint J\cdot E_n^*\,d\tau, \\ H &= A_n i\omega_n H_n \iiint J\cdot E_n^*\,d\tau, \end{aligned} \tag{184}$$

assuming that A_n is large and neglecting all other terms. Substituting in (182) and using (183), we have

$$A_n = -\frac{1}{i\omega_n\iiint gE_n\cdot E_n^*\,d\tau}. \tag{185}$$

Hence, (152) becomes

$$E = \sum \frac{i\omega E_n \iiint J \cdot E_n^* \, d\tau}{\omega^2 - \omega_n^2 - i\omega_n \iiint g E_n \cdot E_n^* \, d\tau}. \tag{186}$$

It is to be remembered that the characteristic functions E_n have been normalized, with the weight function ϵ. The physical significance of the above result is easier to understand if the functions are not normalized; then we should have

$$E = \sum \frac{i\omega E_n \iiint J \cdot E_n^* \, d\tau}{(\omega^2 - \omega_n^2 - i\xi_n\omega_n) \iiint \epsilon E_n \cdot E_n^* \, d\tau} \tag{187}$$

$$\xi_n = \frac{\iiint g E_n \cdot E_n^* \, d\tau}{\iiint \epsilon E_n \cdot E_n^* \, d\tau}. \tag{188}$$

The quantity ξ_n is the dissipated power per unit stored energy.

It can be shown that if the real part of the ratio $E_{\tan}/H_{\tan}$ of the tangential components at the *inner* surface of the resonator is $\mathfrak{R}$, then ξ_n may be expressed as

$$\xi_n = \frac{\iint \mathfrak{R} \, H_{n,\tan} \cdot H_{n,\tan}^* \, dS}{\iiint \epsilon E_n \cdot E_n^* \, d\tau}. \tag{189}$$

This is a very important modification for frequently it is not easy to calculate the exact characteristic functions corresponding to the dissipative case. If the dissipation is small, it is assumed that the characteristic functions in the dielectric are unaltered. This assumption yields the tangential component of H and makes possible the computation of ξ_n.

It will be recalled that in writing (184), all other terms were neglected because A_n was assumed to be large. If ξ_n is small, A_n *is* large. But we should note that if two characteristic values ω_n and ω_{n+1} are nearly equal, then *two adjacent terms* would have large amplitudes, and the above analysis would have to be reexamined and modified.

CHAPTER XIV

Conformal Transformations

1. *Conformal transformations*

Any function,

$$w = f(z), \qquad w = u + iv, \qquad z = x + iy, \tag{1}$$

defines a transformation of points and geometric figures in the complex z-plane into points and figures in the complex w-plane. If w is an analytic (monogenic) function of z, the transformation is *conformal*, except in the vicinity of certain points; the term " conformal " is meant to suggest that

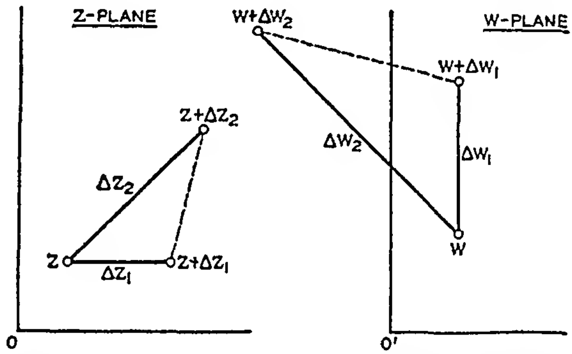

FIG. 14.1. Conformal transformation by means of analytic (monogenic) functions of a complex variable. On an infinitely small scale, shapes are preserved; an infinitely small triangle is transformed into a similar triangle.

on an infinitely small scale the shapes of geometric figures are unaltered by the transformation. An infinitely small triangle in the z-plane is transformed into a similar triangle in the w-plane, Figure 14.1; the size, position and orientation may be changed but not the angles.

To prove the statement we note that, if f is an analytic function,

$$\Delta w = f'(z)\, \Delta z, \tag{2}$$

except for infinitesimals of higher orders. For two displacements from the same point we have

$$\Delta w_1 = f'(z)\, \Delta z_1, \qquad \Delta w_2 = f'(z)\, \Delta z_2. \tag{3}$$

In writing these equations we assume that $f'(z)$ does not become infinite at the point under consideration; if, in addition, $f'(z)$ does not vanish, we can divide the second equation by the first

$$\Delta w_2/\Delta w_1 = \Delta z_2/\Delta z_1. \tag{4}$$

Hence the absolute values of the increments are proportional; and the phase differences, that is the angles between the increments, are equal.

2. *Practical applications*

Many practical applications of analytic functions of a complex variable depend on the fact that the real and imaginary parts of such functions satisfy the two-dimensional Laplace's equation (see Section 5.11). The potential and stream functions in the two-dimensional irrotational flow of liquid satisfy this equation; the electrostatic and magnetostatic potentials satisfy this equation; the stream function of the magnetic field generated by parallel electric current filaments satisfies this equation; etc.

The velocity components ξ, η are obtained either from the *potential function* u,

$$\xi = -\frac{\partial u}{\partial x}, \qquad \eta = -\frac{\partial u}{\partial y} \tag{5}$$

or from the *stream function* v,

$$\xi = -\frac{\partial v}{\partial y}, \qquad \eta = \frac{\partial v}{\partial x}. \tag{6}$$

The function $w = u + iv$ is the *complex potential* and

$$\xi - i\eta = -\frac{dw}{dz}. \tag{7}$$

The speed of flow is given by the absolute value of dw/dz.

In two-dimensional electrostatics the electric intensity is given by a similar expression

$$E_x - iE_y = -\frac{dw}{dz}. \tag{8}$$

Frequent boundary conditions are: (1) the potential is constant along a given boundary, (2) the stream function is constant along a given boundary. Conformal transformations may be used to simplify this boundary; the usual problem is to find a conformal transformation which changes the given boundary into either the real or the imaginary axis.

In the following sections we shall consider several specific transformations, $w = f(z)$, and some field problems which they solve. In the appli-

cations one of the two complex planes is the "coordinate plane," made up of points whose complex potentials are represented by points in the other plane, the "potential plane." Which plane is which, is a matter of choice. Any specific functional relationship between z and w will yield solutions of two different problems, depending on whether we regard w as the complex potential at point z or z as the complex potential at point w.

3. *Function* $w = z^2$

In order to obtain a general idea of the effect of any particular transformation it is usual to select some simple orthogonal network of curves or straight lines and study what happens to it under the transformation. Every transformation is a two-way transformation; thus

$$w = z^2 \tag{9}$$

defines a transformation of the z-plane into the w-plane, while the inverse function, $z = \sqrt{w}$, converts the w-plane into the z-plane. However, it is convenient to consider these transformations separately; two inverse transformations never seem to have quite the same properties.

Substituting $x + iy$ for z and separating the real and imaginary parts, we have

$$x^2 - y^2 = u, \qquad 2xy = v. \tag{10}$$

The straight lines parallel to the u and v-axes are rectangular hyperbolas in the z-plane, Figure 14.2. One family of hyperbolas represents equipotential lines and the other stream lines. According to the convention in the preceding section the solid lines, $u =$ constant, are the equipotential lines and the dashed lines, $v =$ constant, are the stream lines; but the convention is arbitrary. The interchangeability is evident from the transformations $\hat{w} = \pm iw$, $\hat{u} = \mp v$, $\hat{v} = \pm u$.

If the x and y-axes are fixed boundaries of an irrotational flow of liquid, the dotted lines are the stream lines and the solid lines are the velocity potential lines. If the same axes are perfect conductors, at a constant potential but containing electric charge, then the dotted lines are equipotential lines and the solid lines are the lines of electric force.

If z and w are expressed in polar coordinates,

$$z = \rho e^{i\varphi}, \qquad w = re^{i\theta}, \tag{11}$$

then, on substitution in (9), we obtain

$$r = \rho^2, \qquad \theta = 2\varphi. \tag{12}$$

The coordinate lines in the z-plane go into coordinate lines in the w-plane. The upper half of the z-plane, $0 \leq \varphi \leq \pi$, goes into the entire w-plane,

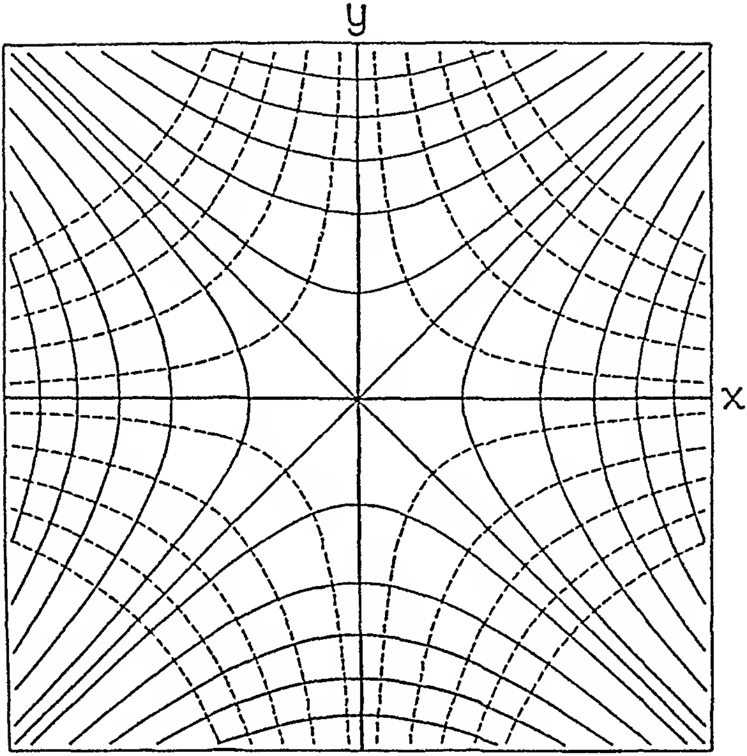

FIG. 14.2. Two families of rectangular hyperbolas which, under the conformal transformation $w = z^2$, become straight lines parallel to the u and v-axes in the w-plane.

$0 \leq \theta \leq 2\pi$; the lower half also goes into the entire w-plane. Points z and $-z$ in the z-plane obviously go into the same point in the w-plane; some distinct geometric figures in the z-plane may go into coincident figures in the w-plane.

4. *Function* $z = \sqrt{w}$

From (12) we have

$$\rho = \sqrt{r}, \qquad \varphi = \tfrac{1}{2}\theta. \tag{13}$$

As the radius from the origin O' in the w-plane makes one complete revolution, the corresponding radius in the z-plane makes only half a revolution. If we think of $\theta = 0$ as referring to the upper side of the positive real axis, Figure 14.3, then $\theta = 2\pi$ would refer to points on the lower half of the axis. The function

$$V = y = \rho \sin \varphi = \sqrt{r} \sin \tfrac{1}{2}\theta \tag{14}$$

satisfies Laplace's equation and vanishes when $\theta = 0, 2\pi$ and nowhere else in the w-plane; hence it is a possible expression for the electric potential in the presence of a conducting half-plane passing through the positive real axis and perpendicular to the w-plane. The potential is finite at finite

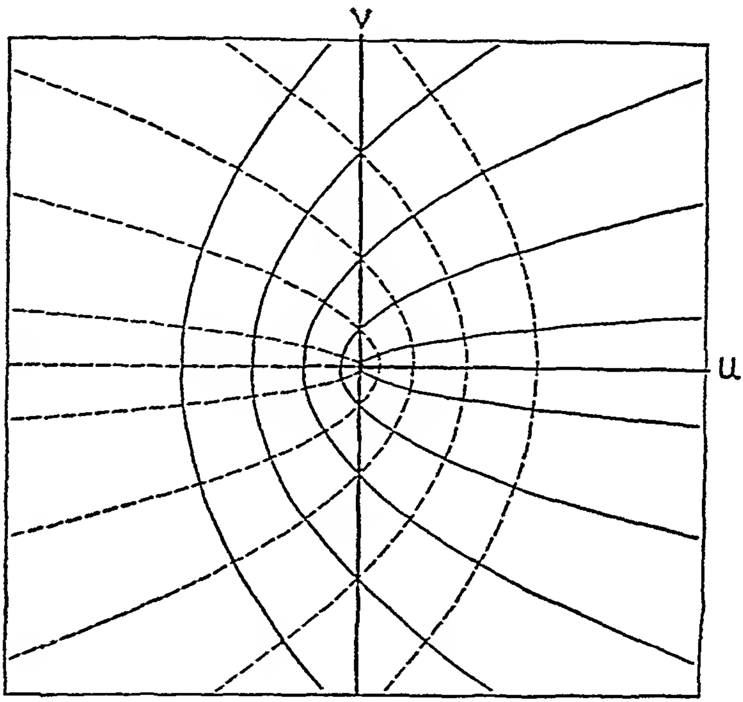

FIG. 14.3. Two families of orthogonal parabolas which represent the cartesian coordinate lines in the z-plane under the transformation $z = \sqrt{w}$.

distances from O'; thus there are no sources parallel to the edge of the conducting plane. Therefore (14) *represents the potential of electric charge* on a conducting half-plane kept at zero potential.

In the present case the parts played by z and w are interchanged; y and x are respectively the potential and stream functions while u and v are the coordinates of a point in the plane. Hence the magnitude of the electric intensity is

$$|E| = |dz/dw| = \tfrac{1}{2}|w^{-1/2}| = \tfrac{1}{2}r^{-1/2} = 1/2\rho. \tag{15}$$

Near the edge of the conducting half-plane, the intensity is large.

In order to find the equipotential and stream lines we eliminate first x

and then y from (10); thus

$$v^2 = 4y^2u + 4y^4, \qquad v^2 = -4x^2u + 4x^4. \tag{16}$$

Both families of curves are parabolas.

So far θ has been restricted to the interval $(0,2\pi)$. If we increase θ from 2π to 4π, the corresponding value of φ increases from π to 2π; that is, going over the w-plane for the second time, we map it on the lower half of the z-plane. To establish a one-to-one correspondence between points in the two planes, we may think of the w-plane as two sheets joined together along the positive u-axis, in such a way that when we cross the junction, called the *branch line* or *cut*, we pass from one sheet to the other. This double sheet is a *Riemann surface* for the representation of the double-valued function $z = \sqrt{w}$. Any straight line or curve from the branch point $w = 0$ to infinity can serve as the branch line or junction of two sheets of the Riemann surface. The interpretation of branch lines as junctions of various sheets on the Riemann surface permits greater freedom in the analysis of functions of a complex variable than the idea of impassable barriers in Section 4.10; but in physical applications the branch lines often appear as barriers: a conducting half-plane in the preceding example, immovable partitions in hydrodynamics, etc.

5. *Function* $w = z^n$

The general power function,

$$w = z^n, \tag{17}$$

possesses properties similar to those of the special functions in the preceding sections. In polar coordinates

$$r = \rho^n, \qquad \theta = n\varphi. \tag{18}$$

The function is single-valued if n is an integer. If n is a fraction, the function is multiple-valued. Thus if φ is increased by 2π (for a given value of ρ), z returns to its original value; but w becomes $\rho^n[\cos(n\varphi + 2n\pi) + i\sin(n\varphi + 2n\pi)]$ and this is, in general, different from the original value. If $n = 1/m$, where m is an integer, then w has m different values and we need m sheets in the Riemann surface for the graphical representation of the function. One complete revolution of the radius from the origin in the z-plane corresponds to $(1/m)$th part of the cycle in the w-plane. If n is an irrational number there are infinitely many values of the function.

With the aid of this function the following question can be answered: what is the potential distribution inside a wedge formed by two conducting half-planes, joined together as in Figure 14.4? Let the boundary of the wedge be $\varphi = 0$ and $\varphi = \psi$. The physical requirement is: the potential

of a conductor must be constant. The field is not affected by the addition of a constant and the potential of any particular conductor may be arbitrarily set equal to zero. The imaginary part of w is

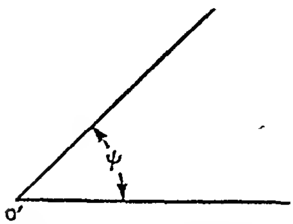

FIG. 14.4. A wedge which is transformed into the upper half of the w-plane by $w=z^{\pi/\psi}$.

$$v = \rho^n \sin n\varphi. \tag{19}$$

It vanishes when $\varphi = 0$. If v is to vanish on the rest of the boundary, $\varphi = \psi$, but not inside the wedge, then

$$n\psi = \pi, \qquad n = \pi/\psi. \tag{20}$$

If $\psi = \pi/2$, we have the case considered in Section 3; $\psi = 2\pi$ yields a "wedge" with the largest possible spread, that is, the exterior of the half-plane of Section 4.

The above solution applies also to the problem of irrotational flow of liquid; the equipotential lines of the electric problem become the stream lines of the hydrodynamical problem.

The magnitude of the electric intensity (or the fluid velocity) is

$$|E| = |dw/dz| = n\,|z^{n-1}| = n\rho^{n-1}. \tag{21}$$

It vanishes near the edge if $n > 1$ or $\psi < \pi$; and it becomes infinite, if $n < 1$ or $\psi > \pi$.

6. *Function* $w = \exp z$

In the case of

$$w = e^z \tag{22}$$

there are simple relationships between the cartesian coordinates in the z-plane and polar coordinates in the w-plane

$$re^{i\theta} = e^{x+iy} = e^x (\cos y + i \sin y), \qquad r = e^x, \qquad \theta = y. \tag{23}$$

The lines $x =$ constant, parallel to the y-axis, become concentric circles in the w-plane; the lines $y =$ constant, parallel to the x-axis, become rays emerging from the origin. A strip of the z-plane, bounded by $y = y_0$ and $y = y_0 + 2\pi$, goes into the entire w-plane.

7. *Function* $z = \log w$

The inverse of (22) is

$$z = \log w; \qquad x = \log r, \qquad y = \theta. \tag{24}$$

This is an infinitely many-valued function since points given by $x = \log r$,

$y = \theta + 2n\pi$, where n is an integer, correspond to the same point in the w-plane.

Let us consider x as the electric potential at the point (r,θ) of the w-plane and see what problem is thereby solved. The electric field is found from (8) (after an interchange of z and w)

$$E_u - iE_v = -\frac{dz}{dw} = -\frac{1}{w} = -\frac{1}{r\exp(i\theta)} = -\frac{\exp(-i\theta)}{r};$$
$$E_u = -\frac{\cos\theta}{r}, \qquad E_v = -\frac{\sin\theta}{r}; \qquad E_r = -\frac{1}{r}, \qquad E_\theta = 0. \tag{25}$$

Equipotential lines are concentric circles; the lines of force are the rays emerging from the center. The center is the " source " of these lines and the problem solved is that of the field of a uniformly charged filament. According to electrostatic laws the integral of ϵE_r, where ϵ is the dielectric constant, round a closed curve surrounding the origin should represent the electric charge q per unit length of the filament. Taking this integral round an equipotential line, we obtain

$$q = -2\pi\epsilon. \tag{26}$$

Next let us regard y as the potential. This makes the potential along each ray constant. Again we have a wedge problem but with this difference: the two half-planes are *at different potentials* and hence must be insulated from each other instead of being connected. The corresponding hydrodynamic problem is the forcing of liquid into or out of a wedge through a narrow slit at the edge.

8. *Function* $z = \cosh w$ *and its inverse*

Next let us consider the following functions

$$w = \cosh^{-1} z \qquad \text{and} \qquad z = \cosh w. \tag{27}$$

The cartesian coordinates in the two planes are related as follows

$$x + iy = \cosh(u + iv) = \cosh u\cos v + i\sinh u\sin v,$$
$$x = \cosh u\cos v, \qquad y = \sinh u\sin v. \tag{28}$$

Dividing the first equation by $\cosh u$, the second by $\sinh u$, squaring and adding, we have an ellipse in the z-plane corresponding to the straight line u = constant in the w-plane. Similarly, v = constant goes into a hyperbola. The equations of the ellipses and hyperbolas are

$$\frac{x^2}{\cosh^2 u} + \frac{y^2}{\sinh^2 u} = 1, \qquad \frac{x^2}{\cos^2 v} - \frac{y^2}{\sin^2 v} = 1. \tag{29}$$

The semimajor and semiminor axes of the ellipses are $\cosh u$ and $\sinh u$; the semi-focal distance is unity. The semi-axes of the hyperbolas are $\cos v$ and $\sin v$; the semi-focal distance is unity. Hence equations (29) represent families of confocal ellipses and hyperbolas.

This transformation may be regarded as a transformation from cartesian to elliptic coordinates (see Chapter 8); indeed, any conformal transformation may be regarded as a transformation from cartesian to orthogonal curvilinear coordinates. The transformation $w = \sqrt{z}$ transforms cartesian into parabolic coordinates.

If u is interpreted as an electric potential, we have the solution for the field between two confocal elliptic cylinders maintained at constant potentials. One of these cylinders may be removed to infinity. As u approaches zero, the cylinder approaches a flat strip.

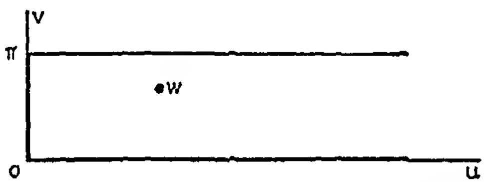

FIG. 14.5. The region bounded by heavy lines corresponds to the upper half of the z-plane if $z = \cosh w$.

Similarly v may be regarded as an electric potential; then confocal hyperbolic cylinders become equipotential surfaces, and may be replaced by conducting surfaces without disturbing the field. As v approaches zero, the hyperbolic cylinder degenerates into a pair of infinite half-planes, passing through the positive and negative real axes from $x = 1$ to $x = \infty$ and from $x = -1$ to $x = -\infty$.

Next let us look at the transformation in the opposite direction, and think of x and y as possible potentials of a point whose *cartesian* coordinates are u and v. Let y be the potential, and let us look for the boundary along which it vanishes, without vanishing inside the region. One such region is bounded by $u = 0$, $v = 0$, and $v = \pi$, Figure 14.5. The potential y vanishes on the boundary and is positive at any point w inside the region. This region corresponds to the upper half of the z-plane for y may take on only positive values and x is unrestricted.

The potential y is sinusoidally distributed along a line parallel to the v-axis and it increases as u increases. The source of the field is at $u = \infty$ and the solution represents the result of "squeezing" the field between two planes. Another application is considered in the next section.

If $u + iv$ is replaced by $\pi(u + iv)/h$ the boundary $v = \pi$ becomes

$v = h$; thus the distance between the parallel planes in Figure 14.5 can be controlled.

9. *Further applications to steady flow*

Let us suppose that

$$w = f(z) \tag{30}$$

transforms the interior of a region (C) into the upper half of the w-plane, Figure 14.6. The boundary of the region goes into the u-axis. Next let us consider the following function

$$W = V + i\Psi = \log \frac{w - w_0}{w - w_0^*} = \log \frac{f(z) - f(z_0)}{f(z) - f(z_0^*)}. \tag{31}$$

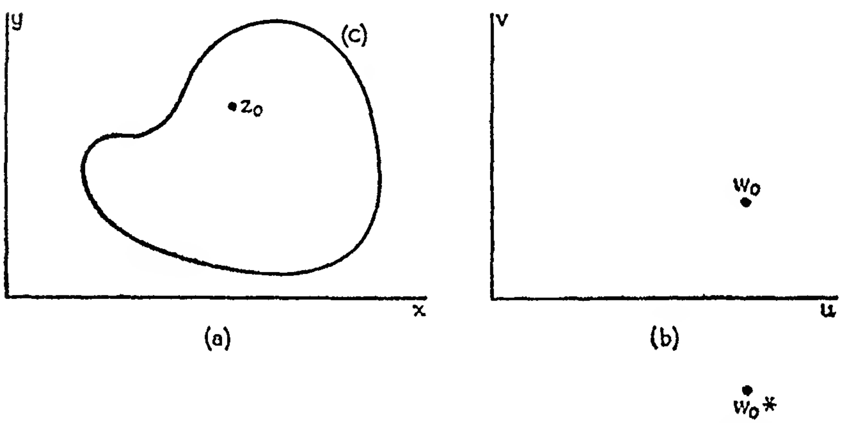

FIG. 14.6. If the region bounded by a closed curve (C) in the z-plane is transformed by $w = f(z)$ into the upper half of the w-plane, then the real part of $\log [f(z) - f(z_0)] - \log [f(z) - f(z_0^*)]$ vanishes on (C) and is logarithmically infinite at $z = z_0$. The transformation represents the solution of two-dimensional flow problems with a source at $z = z_0$ and an equipotential sink on (C).

At $z = z_0$, W is logarithmically infinite. Since

$$V = \log \left| \frac{f(z) - f(z_0)}{f(z) - f(z_0^*)} \right|, \qquad \Psi = ph \frac{f(z) - f(z_0)}{f(z) - f(z_0^*)}, \tag{32}$$

it is the real part V of W that becomes logarithmically infinite at $z = z_0$. In fact, in the immediate vicinity of $z = z_0$, we have $f(z) = f(z_0) + (z - z_0)f'(z_0)$, and therefore

$$W = \log (z - z_0) + \log \frac{f'(z_0)}{f(z_0) - f(z_0^*)}. \tag{33}$$

In Section 7 we have seen that if V is regarded as an electric potential, the logarithmic infinity implies an electric charge at $z = z_0$; if the coefficient of the logarithmic term is unity, the charge is $-2\pi\epsilon$, as in (26). Another interpretation of the logarithmic singularity would be a source or sink of liquid.

The potential V vanishes on the boundary (C) of the region; for, if $w = f(z)$ is on the u-axis in Figure 14.6b, then the absolute values of $(w - w_0)$ and $(w - w_0^*)$ are equal and the logarithm of the ratio is zero. Hence (31) and (32) represent solutions for problems involving a point source; V in (32) is a Green's function for the boundary (C).

The point w_0^* is the optical image of w_0 in the u-axis; by analogy z_0^* is called the image of z_0 in the curved boundary (C) although it is no longer the optical image.

These considerations increase the utility of conformal transformations, since they bring the sources of the field into the picture. In Section 3 it is shown that $w = z^2$ transforms the upper half of the w-plane into the first quadrant of the z-plane; therefore

$$W = \log \frac{z^2 - z_0^2}{z^2 - z_0^{*2}} = \log \frac{(z - z_0)(z + z_0)}{(z - z_0^*)(z + z_0^*)} \tag{34}$$

is the complex potential of a point source in the first quadrant, when the boundaries of the quadrant are at zero potential. Points z_0^* and $-z_0^*$ are the optical images of z_0 in the boundaries; point $-z_0$ is the optical image of these images in the continuations of the boundaries.

The transformation $w = \sqrt{z}$, with (31), solves the case of a source in the presence of a conducting half-plane; thus

$$W = \log \frac{\sqrt{z} - \sqrt{z_0}}{\sqrt{z} - \sqrt{z_0^*}}. \tag{35}$$

In order to obtain the real part of W it is best to express z in polar form; then

$$V = \tfrac{1}{2} \log \frac{\rho - 2\sqrt{\rho\rho_0}\cos \frac{1}{2}(\varphi - \varphi_0) + \rho_0}{\rho - 2\sqrt{\rho\rho_0}\cos \frac{1}{2}(\varphi + \varphi_0) + \rho_0}. \tag{36}$$

The image point $z = z_0^*$ is not on the same sheet of the Riemann surface as $z = z_0$; what we have is

$$\begin{aligned} \sqrt{z_0} &= \sqrt{\rho}\,(\cos \tfrac{1}{2}\varphi_0 + i \sin \tfrac{1}{2}\varphi_0), \\ \sqrt{z_0^*} &= \sqrt{\rho}\,(\cos \tfrac{1}{2}\varphi_0 - i \sin \tfrac{1}{2}\varphi_0), \end{aligned} \tag{37}$$

whereas for $\sqrt{z}$ on the same sheet with $\sqrt{z_0}$ but over $\sqrt{z_0^*}$,

$$\begin{aligned}\sqrt{z} &= \sqrt{\rho}\,[\cos \tfrac{1}{2}(2\pi - \varphi_0) + i \sin \tfrac{1}{2}(2\pi - \varphi_0)] \\ &= \sqrt{\rho}(-\cos \tfrac{1}{2}\varphi_0 + i \sin \tfrac{1}{2}\varphi_0).\end{aligned} \tag{38}$$

Thus the potential is finite at the place which *seems* to be occupied by the image point.

For the boundary in Figure 14.5 we have

$$W = \log \frac{\cosh w - \cosh w_0}{\cosh w - \cosh w_0^*}, \tag{39}$$

where the cartesian coordinates of a point are now u, v. The real part is found as follows. First, we write

$$W = \log \frac{\sinh \frac{1}{2}(w - w_0) \sinh \frac{1}{2}(w + w_0)}{\sinh \frac{1}{2}(w - w_0^*) \sinh \frac{1}{2}(w + w_0^*)}; \tag{40}$$

then we note that the square of the absolute value of sinh ζ is

$$|\sinh \zeta|^2 = \sinh \zeta \sinh \zeta^* = \tfrac{1}{2}[\cosh (\zeta + \zeta^*) - \cosh (\zeta - \zeta^*)] \tag{41}$$

Thus we obtain

$$V = \tfrac{1}{2} \log \frac{[\cosh (u - u_0) - \cos (v - v_0)][\cosh (u + u_0) - \cos (v + v_0)]}{[\cosh (u - u_0) - \cos(v + v_0)][\cosh (u + u_0) - \cos(v - v_0)]}. \tag{42}$$

Choosing

$$W = \log (w - w_0)(w - w_0^*) = \log [f(z) - f(z_0)][f(z) - f(z_0^*)], \tag{43}$$

we find that the normal derivative of V vanishes on (C). The boundary becomes impervious to flow.

10. *Schwartz-Christoffel transformations*

The Schwartz-Christoffel transformation

$$z = A \int^{w} (w - w_1)^{-\vartheta_1/\pi} (w - w_2)^{-\vartheta_2/\pi} \cdots (w - w_n)^{-\vartheta_n/\pi}\, dw + B \tag{44}$$

transforms the region bounded by a polygon in the z-plane into the upper half of the w-plane, Figure 14.7. If we follow the boundary counterclockwise, it is the interior (more generally, the region to the *left* of the boundary) that is transformed into the upper half-plane; the angles ϑ are defined as shown in Figure 14.7 and they are either positive or negative according to whether the rotation is counterclockwise or clockwise. For example, ϑ_1, ϑ_2 and ϑ_3 are positive, but ϑ_4 is negative.

The exterior of the polygon is to the left of the observer moving clock-

in z round the semicircle approaches zero with r_1 and the position of z is unchanged. However, the phase of the integrand is *decreased* by $n_1\pi$ (on account of the clockwise rotation of $w - w_1$ in Figure 14.7), or *increased* by $-n_1\pi$. If the new increments in z make an angle ϑ_1 with the old, then $n_1 = -\vartheta_1/\pi$. Thus as we pass round successive indentations at $w_1, w_2, \cdots$, point z follows a polygon. From the point at infinity on the u-axis we follow the infinite semicircle in the upper half-plane; the phase of the integrand increases by $(n_1 + n_2 + \cdots)\pi = -\vartheta_1 - \vartheta_2 - \cdots = -2\pi$; hence there is no change in the direction followed by z and the path between $z = z_n$ and $z = z_1$ is straight.

The final question is, which of the two regions bounded by the polygon is transformed into the upper half-plane? Let Δw_1 be an increment along the u-axis between two indentations, and Δz_1 the corresponding increment on the polygon between two vertices; the increment $\Delta w_2 = i\Delta w_1$ would take us out into the upper half-plane; by (4) $\Delta z_2 = i\Delta z_1$ will take us to the left of the boundary in the z-plane, that is into the interior of the region bounded by the polygon. Exactly the same rule applies if we follow the polygon clockwise; but then the region on the *left* is the exterior region and the corresponding exterior angles have opposite signs from $\vartheta_1, \vartheta_2 \cdots$ in Figure 14.7.

The complex constants, A and B, control the position, size and orientation of the polygon. Thus B may be so chosen that one of the vertices of the polygon will coincide with some specified point — the origin, for example; A may then be chosen so that one side of the polygon will be of given size and parallel to a given direction.

To illustrate, let us apply the Schwartz-Christoffel transformation to the boundary shown in Figure 14.5. If we follow the boundary counterclockwise and if we choose the points $w = -1$ and $w = 1$ in the potential plane to correspond to the turning points in the coordinate plane (z-plane),

$$z = A\int (w + 1)^{-1/2}(w - 1)^{-1/2}\,dw + B$$

$$= A\int \frac{dw}{\sqrt{w^2 - 1}} + B = A\cosh^{-1} w + B, \tag{47}$$

$$w = \cosh\frac{z - B}{A}.$$

The upper turning point, $z = ih$, is encountered first as we follow the boundary; thus it corresponds to $w = -1$. The lower turning point, $z = 0$, corresponds to $w = 1$. Hence

$$-1 = \cosh\frac{ih - B}{A}, \qquad 1 = \cosh\frac{B}{A}. \tag{48}$$

circles and straight lines into circles and straight lines, we can try to change the given region into a wedge; then we can apply the Schwartz-Christoffel transformation. However, it is easier to work backwards, from the half-plane to the wedge, and then to the given region. Thus, arranging the wedge with respect to the given region as in Figure 14.8, we write

$$w = \hat{z}^{\pi/\psi}, \qquad \bar{z} = 1 + \hat{z}, \qquad z = 1/\bar{z}. \tag{51}$$

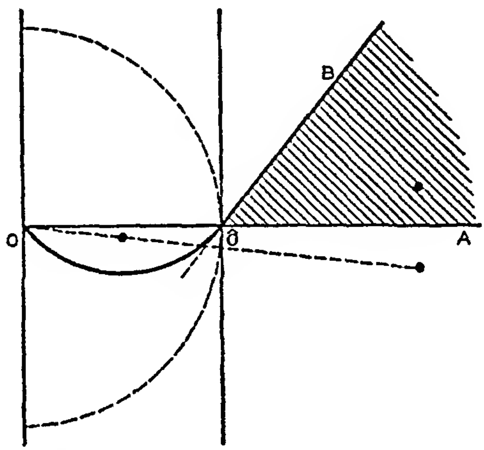

FIG. 14.8. By successive transformations the upper half of the w-plane may be transformed into the region bounded by a straight segment and a circular arc.

The point $\hat{z}$ represents a point within the wedge with $\hat{O}$ as the origin in the complex plane; $\bar{z}$ is the same point if the origin is moved to O; and z is this point after the plane is subjected to the bilinear transformation consisting of taking the reciprocal of $\bar{z}$, assuming that $O\hat{O}$ is unity.

Since the reciprocals of real numbers greater than unity are real numbers smaller than unity, $z = 1/\bar{z}$ transforms $\hat{O}A$ into $\hat{O}O$. The point at infinity goes into O, $\hat{O}$ is unchanged, and $\hat{O}B$ must go into a circle passing through $\hat{O}$ and O. The angle between the arc and segment $O\hat{O}$ must be equal to the wedge angle ψ. Eliminating $\hat{z}$ and $\bar{z}$ from (51), we have

$$w = (z^{-1} - 1)^{\pi/\psi} = (1 - z)^{\pi/\psi} z^{-\pi/\psi}. \tag{52}$$

This example illustrates the possibilities inherent in combining simple known transformations.

Problems

1. Show that the straight segment $O\hat{O}$ corresponds to the positive values of u and the arc to the negative values.

2. Transform the region bounded by two circular arcs into the upper half of the w-plane. Assume that the angle between the arcs is ψ and that α is the angle between one of them and the straight segment connecting the points of intersection of the arcs.
Ans. $w = e^{j\alpha\pi/\psi}(1 - z)^{\pi/\psi} z^{-\pi/\psi}$.

12. *Perturbation of boundaries*

The function $V = \log \rho$ represents the potential of a line filament along $\rho = 0$ of density $q = -2\pi\epsilon$. The equipotential surfaces are cylinders $\rho =$ constant; in particular, the cylinder $\rho = 1$ is at zero potential. By adding another potential function we may alter slightly the shape of the equipotential surfaces in any desired manner. Suppose we wish to calculate the change in the capacitance of a pair of coaxial cylinders due to a slight flattening of the outer conductor so that its shape becomes oval. Since z^2 is an analytic function, its real part $\rho^2 \cos 2\varphi$ is a potential function. This potential function vanishes at $\rho = 0$ and is small in the vicinity. Let us take, therefore,

$$V = \log \rho + k\rho^2 \cos 2\varphi, \tag{53}$$

where k is also small. To obtain the equation of the equipotential $V = 0$, we substitute $\rho = 1 + \delta$, where δ is small, in (53), and neglect all small quantities of order higher than δ

$$\delta + k \cos 2\varphi = 0, \qquad \delta = -k \cos 2\varphi. \tag{54}$$

At $\varphi = 0, \pi$ the change in ρ is negative; at $\varphi = \pi/2, 3\pi/2$ the change is positive. The new equipotential surface is the cylinder $\rho = 1$, slightly flattened in the direction $\varphi = 0$. The minimum and maximum radii are respectively $1 - \delta$ and $1 + \delta$. If the radius of the inner cylinder is small, the effect of the added term on the potential of that cylinder is a small quantity of order higher than δ. Thus for the purpose of calculating the capacitance, the cylinder of slightly oval shape may be replaced by a circular cylinder of average radius. The neglected quantities are of the order of δ^2 and $a^2\delta$, where a is the ratio of the radius of the inner cylinder to the average radius of the outer.

As another example let us take

$$V = \log \rho + k\rho^4 \cos 4\varphi = 0. \tag{55}$$

This boundary has the same type of symmetry as the square since the addition of $90°$ to φ does not alter the equation. If $k > 0$, the effect of the added term is to flatten the cylinder in two perpendicular directions $\varphi = 0$ and $\varphi = 90°$; in each case, the second term is positive and ρ has to be made less than unity in order to make V equal to zero. On the other hand, the cylinder is expanded in the directions $\varphi = 45°$ and $\varphi = 135°$. To draw

the equipotential lines we rewrite (55) as follows,

$$-k \cos 4\varphi = \rho^{-4} \log \rho \equiv f(\rho). \tag{56}$$

Since the absolute value of cos 4φ is never greater than unity, the absolute value of the left side in (56) is never greater than k. The function $f(\rho)$ is a large negative quantity near $\rho = 0$, becomes zero for $\rho = 1$, reaches the maximum value 0.09197 when $\rho = 1.284 \cdots$, and vanishes when $\rho = \infty$. This means that if $k > 0.09197$, no value of ρ can correspond to $\varphi = 45°$. Starting with $\varphi = 0$ to trace the curve $V = 0$, we slide upward on the curve

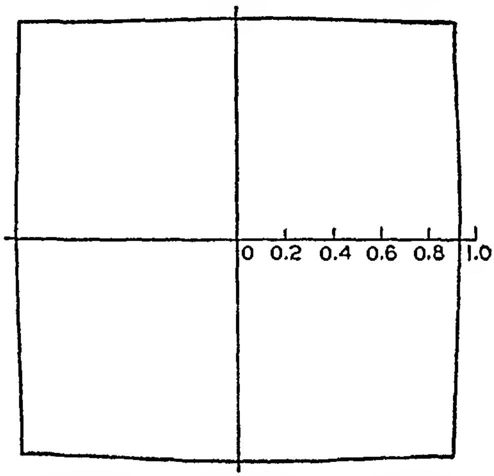

FIG. 14.9. Some potential problems may be solved by a perturbation of boundaries; thus when $k = 0.09197$ the second term in equation (55) distorts one of the original circular equipotential lines into an almost perfect square, without appreciably altering the shape of the equipotential lines near the center.

$f(\rho)$ and find that ρ increases. As soon as we reach $\varphi = \varphi_0$, where φ_0 is defined by the eqnation, $-k \cos 4\varphi_0 = 0.09197$, we shall find ourselves at the maximum point of the curve $f(\rho)$, and a further increase in ρ will cause φ to diminish. The boundary $V = 0$ will not close around the origin as it does when $k < 0.09197$. When $k = 0.09197$, the equipotential $V = 0$ looks as shown in Figure 14.9. The curve is not quite a square. The side of the pseudo-square is 1.864 and the diagonal 2.568; the ratio of the " diagonal " to the " side " is 1.38 instead of 1.41 $\cdots$ for the true square.

Let a be the ratio of the radius of the inner cylinder to the radius, $\rho = 1$, of the equipotential for the case $k = 0$. The capacitance per unit length of the cylinders is

$$C = \frac{-2\pi\epsilon}{\log a - \log 1} = \frac{2\pi\epsilon}{\log (1/a)}. \tag{57}$$

The potential

$$V = \log \rho + 0.09197\rho^4 \cos 4\varphi \tag{58}$$

is zero on our approximate square; but the potential of the cylinder $\rho = a$ remains practically the same as before even if a is as large as 0.5. Hence, (57) represents the approximate capacitance of a circular cylinder inside a square cylinder. Since the side of the square is $b = 1.864$, (57) becomes

$$C = \frac{2\pi\epsilon}{\log (b/2a) + 0.0704} \tag{59}$$

This capacitance is practically equal to the capacitance of the circular cylinder inscribed in the square cylinder in the presence of the same inner conductor.

References

1. R. Rothe, F. Ollendorff and K. Pohlhausen, *Theory of Functions as Applied to Engineering Problems*, Technology Press, Massachusetts Institute of Technology, Cambridge, 1942.
2. A. S. Ramsey, *A Treatise on Hydromechanics*, Part II, *Hydrodynamics*, G. Bell and Sons, Ltd., London, 1913.

CHAPTER XV

Contour Integration

1. *Contour integration*

The integral of a function $f(z)$ is defined as the limit of the sum of products $f(z_m)\Delta z_m$, in which the values of the function and successive increments are taken along a given curve connecting two given points (see equation 6–22). Naturally, the integral may be expected to depend on the integrand and on the path, or contour, of integration; and yet, under some conditions, the integral is independent of the latter.

Some possibilities are suggested by the examples in Section 6.6. Broadly the situation may be summarized as follows:

(a) If the antiderivative of $f(z)$ does not exist, the integral depends on the path of integration as well as on its end points. This happens when the real and imaginary parts of $f(z)$ do not satisfy the Cauchy-Riemann conditions (5–41). In this case there is no alternative to the direct evaluation of the line integrals of the real and imaginary parts of the integrand. In applied mathematics this is not an important case.

(b) If the antiderivative $F(z)$ of $f(z)$ exists and is a single-valued function, the integral is independent of the path of integration and is equal to $F(z_2) - F(z_1)$, where z_1, z_2 are the end points of the path. This happens when the Cauchy-Riemann conditions are satisfied by the real and imaginary parts of $f(z)$ in the entire complex plane.

(c) If the antiderivative is a multiple-valued function, the integral may or may not depend on the path of integration. The integral is still equal to $F(z_2) - F(z_1)$ provided the proper values of $F(z)$ are taken at the end points. The proper value at the starting point $z = z_1$ is obtained from $dF(z) = f(z)\,dz$; the initial value of $f(z)$ should be given. The final value at $z = z_2$ is obtained by following $F(z)$ continuously along the path of integration.

The second and third possibilities occur when the integrand is a monogenic or analytic function except at some isolated points. If the antiderivative is actually known, there is nothing further to add; but certain integrals of analytic functions can be evaluated without knowing the antiderivatives. The present chapter is devoted to these methods.

Problems

1. Evaluate the following integral: $\int_C x\,dz$, where C is the straight line joining the origin to the point $z = a + ib$. *Ans.* $\frac{1}{2}a(a + ib)$.

2. Evaluate the integral in Problem 1 between the same points but following the straight path from $z = 0$ to $z = a$, and then from $z = a$ to $z = a + ib$.

Ans. $\frac{1}{2}a(a + 2ib)$.

3. Evaluate $\int \sqrt{z}\,dz$ from $z = 1$ to $z = a + ib$ along the straight path connecting these points, or along any path which can be obtained from the straight path by a continuous deformation which keeps the end points fixed and does not cross the origin. Assume that the initial value of the integrand is unity.

Ans. $\frac{2}{3}(a + ib)\sqrt{a + ib} - \frac{2}{3}$, where $\sqrt{a + ib} = |\sqrt{a^2 + b^2}| \exp [i \tan^{-1} (b/a)]$ and $0 < \tan^{-1} (b/a) < \pi$ when $b > 0$, whereas $-\pi < \tan^{-1} (b/a) < 0$ when $b < 0$. When $b = 0$, the integral along the straight path is not defined if $a < 0$, since *in passing the branch point $z = 0$ we lose track of the particular value to be* assigned to the integrand.

2. *Cauchy's theorem*

If $f(z)$ is analytic at every point of the region bounded by a closed contour C, then

$$\int_C f(z)\,dz = 0. \tag{1}$$

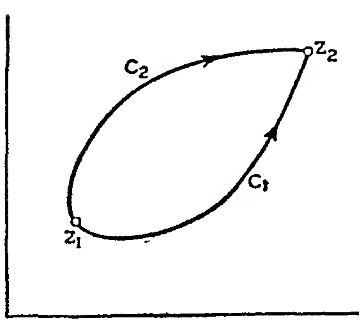

FIG. 15.1. Under certain conditions the contour integral of $f(z)$ depends only on the end points, z_1 and z_2, and not on the path of integration.

The obvious corollary is, Figure 15.1,

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz, \tag{2}$$

provided $f(z)$ is analytic in the region bounded by C_1 and C_2.

If u and v are the real and imaginary parts of $f(z)$,

$$\int_C f(z)\,dz = \int_C (u + iv)(dx + i\,dy)$$
$$= \int_C (u\,dx - v\,dy) + i\int_C (v\,dx + u\,dy). \tag{3}$$

By Green's theorem, equations (6–31) and (6–32), the line integrals may be transformed into surface integrals over the area bounded by C; thus

$$\int_C f(z)\,dz = \iint \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) dx\,dy + i\iint \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy. \tag{4}$$

If the function is analytic at *every* point in the region, the integrands vanish in consequence of the Cauchy-Riemann equations (5–41), and the theorem is proved.

Analyticity at *every* point in the region is a sufficient but not a necessary condition for (1). For instance, neither of the following functions

$$f_1(z) = \frac{1}{z} = \frac{x - iy}{x^2 + y^2},$$
$$f_2(z) = \frac{1}{z^2} = \frac{(x - iy)^2}{(x^2 + y^2)^2} = \frac{(x^2 - y^2) - 2ixy}{(x^2 + y^2)^2}, \tag{5}$$

is analytic at $z = 0$; but the integral of $f_2(z)$ round a closed contour surrounding the origin vanishes while the integral of $f_1(z)$ does not. The reader should note very carefully that if he were to take the partial derivatives of the real and imaginary parts of the above functions and substitute in (4), he should find that to all *appearances* the integrands vanish everywhere; but in fact neither u nor v, nor their derivatives, possess any value when $x = y = 0$.

The proofs of Green's and Cauchy's theorems are based on a connected curve (C). Annular regions such as that shown in Figure 15.2 have two separate boundaries, the inner and the outer, and a question arises as to the relative directions of integration round the two boundaries. Of course, if $f(z)$ is analytic in the region bounded by the external contour, then Cauchy's theorem would apply to each boundary separately, and the total integral round the entire periphery would vanish regardless of the relative directions followed along each closed curve. Suppose, however, that at some point O within the region bounded by the inner contour, $f(z)$ fails to be analytic, whereas in the region between the two contours $f(z)$ is analytic everywhere. Cauchy's theorem would not apply to each contour separately; but it would apply to the combined contour, provided we follow its parts in proper directions, and provided $f(z)$ is *single-valued*. Now this condition of single-valuedness is automatically satisfied if C is a single curve and if $f(z)$ is analytic in the interior; but in an annular region $f(z)$ may be analytic without being single-valued. The function $f(z) = \sqrt{z}$ is an example; it is analytic in the annulus and yet multiple-valued.

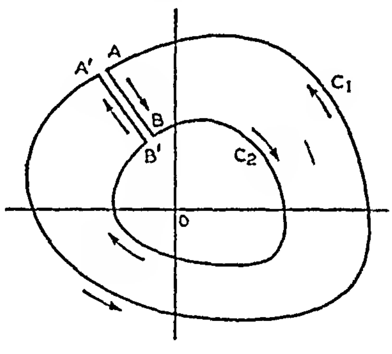

FIG. 15.2. Transformation of an annular region into a simply connected region by means of a cut.

The extension of Cauchy's theorem to annular regions depends on their conversion into regions bounded by a single curve, as illustrated in Figure 15.2. We cut the annulus by two infinitely close lines AB and $A'B'$. Cauchy's theorem applies to the connected contour $(C_1, AB, C_2, B'A')$ if $f(z)$ is analytic inside the cut annulus. If $f(z)$ is also single-valued, the integrals along AB and $B'A'$ cancel each other, since the increments in $f(z)$ will be equal in magnitude but opposite in sign. Hence,

$$\int_{C_1} f(z)\,dz + \int_{C_2} f(z)\,dz = 0. \tag{6}$$

If $-C_2$ represents C_2 in the opposite direction,

$$\int_{C_1} f(z)\,dz = \int_{-C_2} f(z)\,dz; \tag{7}$$

that is, a *closed contour can shrink or expand without changing the integral of a single-valued function if in this deformation the contour does not pass through a point in which $f(z)$ fails to be analytic.*

The region bounded by a single closed curve which does not intersect itself, as in Figure 15.1, is called a *simply-connected region.* Any closed curve in this region can shrink to a point without crossing the boundary. The annulus in Figure 15.2 is a *doubly-connected* region which can be converted into a simply-connected region by one cut. A triply-connected region is one which can be converted into a simply-connected region by two cuts; etc. The extension of Cauchy's theorem to multiply-connected regions depends on these cuts, which serve to define the proper direction in which each separate closed contour is to be followed.

Closed curves which intersect themselves as in a figure of eight are best treated as several separate contours.

3. *Integration of $f(z) = (z - z_0)^n$*

If n is a positive integer, $f(z) = (z - z_0)^n$ is analytic in the finite part of the plane, and by Cauchy's theorem it's integral round a closed contour is zero. If n is a negative integer, $f(z)$ is single-valued but becomes infinite at $z = z_0$; this is the only point where $f(z)$ fails to be analytic. By Cauchy's theorem the integral will vanish if the contour does not enclose $z = z_0$; otherwise the contour may be deformed into a circle with its center at $z = z_0$, Figure 15.3.

On this circle

$$z - z_0 = re^{i\theta}, \qquad dz = ire^{i\theta}\,d\theta, \tag{8}$$

and

$$\int_C (z - z_0)^n\,dz = ir^{n+1}\int_0^{2\pi} e^{i(n+1)\theta}\,d\theta = \frac{r^{n+1}}{n+1}\,e^{i(n+1)\theta}\Big|_0^{2\pi}, \tag{9}$$

provided $n \neq -1$. For integral values of n the exponential function is unity at $\theta = 0$ and $\theta = 2\pi$; hence the integral vanishes. If $n = -1$,

$$\int_C \frac{dz}{z - z_0} = i\int_0^{2\pi} d\theta = 2\pi i. \tag{10}$$

To summarize: the integral of $(z - z_0)^n$ vanishes for all integral values of n if the contour does not enclose $z = z_0$; if the contour encloses $z = z_0$,

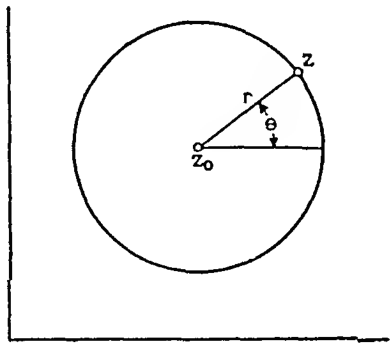

FIG. 15.3. *An illustration for the contour integration of $f(z) = (z - z_0)^n$.*

the integral vanishes for all integral values except $n = -1$ when it is equal to $2\pi i$. If n is a negative integer, the contour cannot pass through $z = z_0$, since the integral does not exist there.

If n is a fraction, the exponential function in (9) does not return to its original value after going round the circle. In this case the integral depends, not only on the radius of the circle, but also on the starting point $\theta = \theta_0$, for

$$\int_C (z - z_0)^n \, dz = \frac{r^{n+1}}{n+1}\left[e^{i2(n+1)\pi} - 1\right] e^{i\theta_0}. \tag{11}$$

In the case of multiple-valued integrands the contours of integration may be subjected only to those deformations which keep the end points fixed, and which do not involve passing through a singularity of the integrand.

4. *Cauchy's integral formulas*

If $f(z)$ is analytic in the interior of C and on C itself, then

$$f(z) = \frac{1}{2\pi i}\oint \frac{f(t)}{t - z}\,dt \tag{12}$$

for any point z in the interior of C. The directed circle over the integral sign is the usual sign for a counterclockwise integral round a closed curve.

Under the stated conditions we may let the contour shrink into an infinitesimal circle centered at $t = z$. On this circle

$$f(t) - f(z) = f'(z)(t - z) + \epsilon(t - z), \tag{13}$$

where ϵ approaches zero with $t - z$. Therefore

$$\oint \frac{f(t)}{t - z}\,dt = \oint \frac{f(z)}{t - z}\,dt + \oint f'(z)\,dt + \oint \epsilon\,dt. \tag{14}$$

The numerator of the first integral on the right is independent of the variable of integration and can be taken outside the integral sign; by (10) the integral is $2\pi i f(z)$. The second integral vanishes in accordance with (9); the last integral also vanishes since ϵ approaches zero with the radius of the circle. Thus we have (12).

The successive derivatives of (12) are

$$\begin{aligned} f'(z) &= \frac{1}{2\pi i}\oint \frac{f(t)}{(t - z)^2}\,dt, \qquad f''(z) = \frac{1.2}{2\pi i}\oint \frac{f(t)}{(t - z)^3}\,dt, \ldots, \\ f^{(n)}(z) &= \frac{n!}{2\pi i}\oint \frac{f(t)}{(t - z)^{n+1}}\,dt. \end{aligned} \tag{15}$$

Formally these equations are obtained by successive differentiations under the integral sign. But we have not proved that the derivative of an integral is the integral of a derivative; and it is conceivable that an interchange of two limiting processes may lead to different answers. The only safe procedure is to base the argument on the definition of the derivative. Thus

$$f(z + \Delta z) = \frac{1}{2\pi i}\oint \frac{f(t)}{t - z - \Delta z}\,dt; \tag{16}$$

subtracting (12), we have

$$f(z + \Delta z) - f(z) = \frac{1}{2\pi i}\oint \frac{f(t)\Delta z}{(t - z)(t - z - \Delta z)}\,dt; \tag{17}$$

and therefore

$$f'(z) = \lim \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim \frac{1}{2\pi i}\oint \frac{f(t)\,dt}{(t - z)(t - z - \Delta z)}. \tag{18}$$

To show that as Δz approaches zero we obtain the first integral in (15), we take the difference

$$\oint \frac{f(t)\,dt}{(t - z)(t - z - \Delta z)} - \oint \frac{f(t)\,dt}{(t - z)^2} = \oint \frac{f(t)\,\Delta z\,dt}{(t - z)^2(t - z - \Delta z)}, \tag{19}$$

and prove that it vanishes with Δz. Now the absolute value of the sum of two complex numbers is never greater than the sum of the absolute values of the numbers. Since $|dt|$ is ds, the differential element of length, we have

$$\left|\oint F(t)\,dt\right| \leq \oint |F(t)|\,|dt| \leq Ml, \tag{20}$$

where M is the largest absolute value of $F(t)$ and l is the length of the contour. When applying this inequality to (19), we shall have to assume that $|f(t)|$ possesses an upper bound on C. For any point z inside the contour, Figure 15.4, the line $|t - z|$ will have a minimum length for some point on C; thus, $M = |f(t)|_{\max} R_{\min}^{-3}\,|\Delta z|$, where $R_{\min}$ is the least distance between the points on C and on a circle of radius $|\Delta z|$ centered on $t = z$. Since M vanishes with $|\Delta z|$, we have proved the first equation in (15).

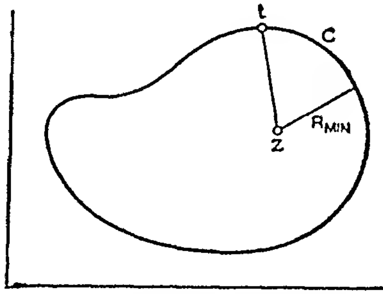

FIG. 15.4. The distance $|t - z|$ between a point on the contour C and an interior point has a minimum value $R_{\min}$.

The same argument applies to all higher derivatives of the contour integral.

5. *Taylor's series*

Maclaurin's and Taylor's series may be obtained very readily from Cauchy's integral formulas. Let C be a circle centered at the origin, and let $f(z)$ be analytic on C and in the interior, so that we can use (12). For any point z in the interior

$$\frac{1}{t - z} = \frac{1}{t\left(1 - \dfrac{z}{t}\right)} = \frac{1}{t} + \frac{z}{t^2} + \frac{z^2}{t^3} + \cdots + \frac{z^{n-1}}{t^n} + \frac{z^n}{(t - z)t^n}. \tag{21}$$

Substituting in (12) and applying (15) with $z = 0$, we have

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\oint \frac{f(t)}{t}\,dt + \frac{z}{2\pi i}\oint \frac{f(t)}{t^2}\,dt + \cdots + z^n \oint \frac{f(t)\,dt}{(t - z)t^n} \\ &= f(0) + zf'(0) + \frac{z^2}{2!}f''(0) + \cdots + \frac{z^{n-1}}{(n - 1)!}f^{(n-1)}(0) \\ &\qquad + z^n \oint \frac{f(t)\,dt}{(t - z)t^n}. \end{aligned} \tag{22}$$

Similarly from

$$\frac{1}{t-z} = \frac{1}{(t-a)-(z-a)} = \frac{1}{t-a} + \frac{z-a}{(t-a)^2} + \cdots \tag{23}$$

and Cauchy's integral formulas we obtain Taylor's series.

If R is the radius of the circle and M the maximum absolute value of $f(t)$ on this circle, the absolute value of the remainder in (22) is

$$\left| z^n \oint \frac{f(t)\,dt}{(t-z)t^n} \right| \leq |z|^n \frac{2\pi RM}{R^n \, |t-z|_{\min}}. \tag{24}$$

If $|z| < R$, that is if z is inside the circle, then the remainder approaches zero as n increases indefinitely, and the series is convergent.

Thus we have shown that the analyticity of a function in a circle centered at a given point is a sufficient condition for the convergence of the power series expansion about that point. Since every convergent power series represents automatically an analytic function, the analyticity of the function is a necessary condition for the existence of the power series. The function $\exp(-1/z^2)$ cannot be expanded in a convergent series of positive integral powers of z because it is not analytic at $z = 0$, that is, at a point which is automatically inside every circle centered on it.

6. *Laurent's series*

Consider the following function

$$f(z) = \frac{3-2z}{(1-z)(2-z)} = \frac{1}{1-z} + \frac{1}{2-z}. \tag{25}$$

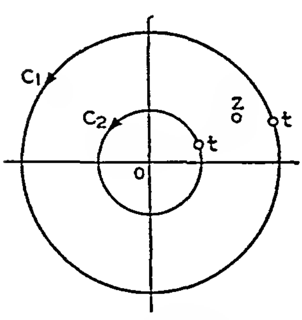

FIG. 15.5. Concentric circles used in obtaining Laurent's expansion.

If $|z| < 1$, each term can be expanded as a geometric series of positive powers of z; if $|z| > 2$, both terms can be expanded in series of negative powers of z; but in the intermediate region one term can be expanded only in the series of positive powers and the other in the series of negative powers. Hence in the annular ring bounded by the circles $|z| = 1$ and $|z| = 2$, $f(z)$ can be expanded only in a series which includes both positive *and* negative powers of z. This is an example of Laurent's series.

We shall now prove the following general theorem. *Let $f(z)$ be single-valued and analytic in an annular ring centered at $z = 0$,* Figure 15.5; *then*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n, \tag{26}$$

where

$$a_n = \frac{1}{2\pi i}\oint_{C_1} t^{-n-1}f(t)\,dt, \qquad n \geq 0;$$
$$a_{-n} = \frac{1}{2\pi i}\oint_{C_2} t^{n-1}f(t)\,dt, \qquad n > 0. \tag{27}$$

From Cauchy's integral formula we have

$$f(z) = \frac{1}{2\pi i}\oint_{C_1}\frac{f(t)}{t-z}\,dt - \frac{1}{2\pi i}\oint_{C_2}\frac{f(t)}{t-z}\,dt. \tag{28}$$

On C_1 we can expand $1/(t-z)$ in a series of positive powers of z/t, and on C_2 in a series of negative powers

$$f(z) = \frac{1}{2\pi i}\oint_{C_1}\left(\frac{1}{t}+\frac{z}{t^2}+\frac{z^2}{t^3}+\cdots\right)f(t)\,dt$$
$$+\frac{1}{2\pi i}\oint_{C_2}\left(\frac{1}{z}+\frac{t}{z^2}+\frac{t^2}{z^3}+\cdots\right)f(t)\,dt. \tag{29}$$

Thus we have the required expansion. The proof that the remainder term of the second series ultimately vanishes is identical with the corresponding proof in the preceding section.

If the ring is centered at $z = a$,

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z-a)^n,$$

$$a_n = \frac{1}{2\pi i}\oint_{C_1}(t-a)^{-n-1}f(t)\,dt, \qquad n = 0,1,2,\cdots, \tag{30}$$

$$a_{-n} = \frac{1}{2\pi i}\oint_{C_2}(t-a)^{n-1}f(t)\,dt, \qquad n = 1,2,3,\cdots.$$

This may be proved either by using expansions analogous to (23), or from (27) by translating the origin in the complex plane to $z = a$.

7. *The theorem of residues*

The coefficient a_{-1} in Laurent's expansion of $f(z)$,

$$f(z) = \cdots + a_{-2}(z-a)^{-2} + a_{-1}(z-a)^{-1} + a_0 + a_1(z-a) + \cdots \tag{31}$$

is called the *residue of $f(z)$ at $z = a$*; it is denoted as follows

$$a_{-1} = \operatorname{Res} f(z)_{z=a}. \tag{32}$$

From the integrals in Section 3 we have

$$\oint f(z)\,dz = 2\pi i a_{-1} = 2\pi i \operatorname{Res} f(z)_{z=a}, \tag{33}$$

for any contour enclosing $z = a$ but no other singularity of $f(z)$.

If the contour encloses several singularities, we can let it shrink to enclose each singularity separately, Figure 15.6. The contributions to the integral from each pair of straight-line segments traversed in opposite directions cancel each other, and the integral equals the sum of the residues multiplied by $2\pi i$,

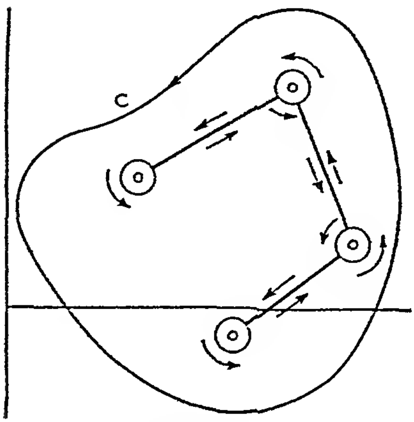

Fig. 15.6. Deformation of a contour enclosing several poles.

$$\oint f(z)\,dz = 2\pi i \sum \operatorname{Res} f(z). \tag{34}$$

If Laurent's expansion of $f(z)$ is known in the vicinity of each singularity, then the residues are found by inspection. Otherwise we proceed as follows. If

$$\lim\,(z - a)^n f(z) \text{ as } z \to a \tag{35}$$

exists and is different from zero for some positive integer n, the point $z = a$ is said to be a *pole of order n*. For a pole of order n, Laurent's series begins with $a_{-n}(z - a)^{-n}$. *Poles of the first order are called simple poles*. If $n = \infty$, $z = a$ is said to be an *essential singularity*. For example, $z = 0$ is an essential singularity of $\exp(1/z)$. To obtain the residue at a pole $z = a$, we determine the order n of the pole; that is, the smallest value of the exponent for which the limit in (35) exists and is different from zero. If $n = 1$, the limit is the residue

$$\operatorname{Res} f(z)_{z=a} = \lim\,(z - a)f(z) \text{ as } z \to a. \tag{36}$$

Otherwise we let a_{-n} denote the limit and obtain

$$a_{-n+1} = \lim\,(z - a)^{n-1}[f(z) - a_{-n}(z - a)^{-n}]. \tag{37}$$

If $n = 2$, this is the residue and the process is terminated; otherwise we calculate

$$a_{-n+2} = \lim\,(z - a)^{n-2}[f(z) - a_{-n}(z - a)^{-n} - a_{-n+1}(z - a)^{-n+1}], \tag{38}$$

and continue until we find a_{-1}.

Consider, for example,

$$f(z) = \frac{1}{1 - \cos z}. \tag{39}$$

The point $z = 0$ is a singularity. Recalling the following rule of calculus

$$\lim_{z \to a} \frac{N(z)}{D(z)} = \frac{N^{(n)}(a)}{D^{(n)}(a)}, \tag{40}$$

where n is the lowest order for which the derivatives of the numerator and denominator do not vanish simultaneously, we have

$$\begin{aligned} &\lim_{z\to 0} \frac{z}{1 - \cos z} = \frac{1}{\sin z\,|_{z=0}} = \infty, \\ &\lim \frac{z^2}{1 - \cos z} = \frac{2}{\cos z\,|_{z=0}} = 2 = a_{-2}. \end{aligned} \tag{41}$$

$$\begin{aligned} a_{-1} &= \lim z\left(\frac{1}{1 - \cos z} - \frac{2}{z^2}\right) = \lim \frac{z^2 - 2 + 2\cos z}{z(1 - \cos z)} \\ &= \lim \frac{2z - 2\sin z}{1 - \cos z + z \sin z} = \lim \frac{2 - 2\cos z}{2 \sin z + z \cos z} \\ &= \lim \frac{2 \sin z}{3 \cos z - z \sin z} = 0. \end{aligned} \tag{42}$$

Consequently

$$\oint \frac{dz}{1 - \cos z} = 0, \tag{43}$$

for any contour which does not enclose poles of the integrand other than $z = 0$. Since these other poles are at $z = \pm 2n\pi$, $n = 1, 2, 3, \cdots$, the contour can be a circle centered at $z = 0$ if the radius is smaller than 2π.

In many cases such as the above, the analysis can be simplified by expanding the functions directly in power series

$$\begin{aligned} \frac{1}{1 - \cos z} &= \frac{1}{1 - 1 + \frac{1}{2}z^2 - \frac{1}{24}z^4 + \cdots} = \frac{2}{z^2(1 - \frac{1}{12}z^2 + \cdots)} \\ &= \frac{2(1 + \frac{1}{12}z^2 + \cdots)}{z^2} = \frac{2}{z^2} + \frac{1}{6} + \cdots. \end{aligned} \tag{44}$$

The expansion should be carried out only far enough to include the residue.

The following theorem is frequently useful. If

$$f(z) = \frac{N(z)}{D(z)}, \tag{45}$$

and $z = a$ is a simple zero of $D(z)$, while $N(a)$ is finite and different from zero,

$$\operatorname{Res} f(z) = \frac{N(a)}{D'(a)}. \tag{46}$$

The proof is left to the reader. [*Hint:* Use the following theorem from calculus: $\lim (uv) = \lim (u) \lim (v)$.]

Problems

1. Find the poles and residues of $f(z) = 1/(z^2 + pz + q)$. *Ans.* The poles are $z_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{1}{4}p^2 - q}$; the residues are respectively $(z_1 - z_2)^{-1} = (\frac{1}{4}p^2 - q)^{-1/2}$ and $(z_2 - z_1)^{-1} = -(\frac{1}{4}p^2 - q)^{-1/2}$.

2. Find the poles and residues of $z/(z^2 - 3z + 2)$.
Ans. $z = 1, 2$; residues $= -1, 2$.

3. Find the poles and residues of $1/(z^4 - 1)$.
Ans. $z = 1, i, -1, -i$; residues $= \frac{1}{4}, \frac{1}{4}i, -\frac{1}{4}, -\frac{1}{4}i$.

4. Find the poles and residues of $1/(z^n - 1)$.
Ans. $z_m = \exp(2im\pi/n)$, where $m = 0, 1, 2, \cdots n - 1$; residues $= \frac{1}{n} z_m$.

5. Find the poles and residues of $\cot z$.
Ans. $z_{\pm n} = n\pi$, $n = 0, 1, 2, \cdots$; residues $= 1$.

6. Find the poles and residues of $\csc z$.
Ans. $z_{\pm n} = n\pi$, $n = 0, 1, 2, \cdots$; residues $= (-)^n$.

7. Find $\oint \frac{z\,dz}{z^2 - 3z + 2}$. *Ans.* $-2\pi i$ if the contour encloses $z = 1$ but not $z = 2$; $4\pi i$ if it encloses $z = 2$ but not $z = 1$; $2\pi i$ if it encloses both points; and 0 if it encloses neither.

8. Find the residue of $z^4/(z - 1)^2$. *Ans.* 4.

9. Find the residues of $(z - 1)/(z + 1)(z + 2)$ at $z = -1$ and $z = -2$.
Ans. -2 and 3.

10. Find the residues of $z^2/(z - 1)(z + 1)^2$ at $z = 1$ and $z = -1$. *Ans.* $\frac{1}{4}$ and $\frac{3}{4}$.

11. Calculate $\oint \frac{z^2\,dz}{i(z - 1)(z + 1)^2}$ round the circle of radius $\sqrt{2}$ with center at $z = 0$. *Ans.* 2π.

12. Calculate the residue of $e^z z^{-5}$ at $z = 0$. *Ans.* $\frac{1}{24}$.
Hint: The answer can be found very simply.

13. Obtain the residues of $z^{-3} \sin z$ and $z^{-3} \cos z$ at $z = 0$. *Ans.* 0, $-\frac{1}{2}$.

14. Find the residue of $(z - 2\pi)^{-4} \sin z$ at $z = 2\pi$. *Ans.* $-\frac{1}{6}$.

8. *An interpretation of the theorem of residues*

The real and imaginary parts of an analytic function of a complex variable satisfy the two dimensional Laplace's equation, and thus may represent the following physical functions: the potential (electric, magnetic, velocity), the flux function, the cartesian components of the velocity in an irrotational flow, the cartesian components of the electric current density (in a steady flow), etc. The poles of functions of a complex variable correspond to physical "sources" or "sinks," and the residues are related to the magnitudes of these sources.

For instance, if a liquid is flowing out of a point, uniformly in all directions, and if I is the emission rate, the polar components of the density of flow are

$$v_\rho = I/2\pi\rho, \qquad v_\varphi = 0. \tag{47}$$

From this we obtain the cartesian components

$$v_x = v_\rho \cos\varphi = \frac{I\cos\varphi}{2\pi\rho}, \qquad v_y = \frac{I\sin\varphi}{2\pi\rho}. \tag{48}$$

We now combine them into an analytic function

$$v_x - iv_y = \frac{I(\cos\varphi - i\sin\varphi)}{2\pi\rho} = \frac{I}{2\pi z}, \tag{49}$$

where $z = \rho\exp(i\varphi) = x + iy$. Thus a simple source at $z = 0$ corresponds to a simple pole of the "complex velocity function." The residue, $I/2\pi$, is very simply related to the emission rate.

Next, consider two equal and opposite poles, that is, a source and a sink, at $z = l/2$ and $z = -l/2$; then

$$v_x - iv_y = \frac{I}{2\pi(z - \frac{1}{2}l)} - \frac{I}{2\pi(z + \frac{1}{2}l)} = \frac{Il}{2\pi(z^2 - \frac{1}{4}l^2)}. \tag{50}$$

Let $l \to 0$ and $I \to \infty$ in such a way that the "moment," $p = Il$, of the "dipole" or "doublet" remains constant; then

$$v_x - iv_y = \frac{p}{2\pi z^2}. \tag{51}$$

This function has a pole of the second order at $z = 0$.

A source of the nth order is similarly formed by n simple sources of equal intensity, equispaced on the circumference of an infinitesimal circle, and a simple sink of n times the intensity of each circumferential source. The complex velocity is found to be proportional to z^{-n}, a function having a pole of order n at $z = 0$.

Consider now the integral

$$\oint f(z)\,dz = \oint (v_x - iv_y)(dx + i\,dy)$$
$$= \oint (v_x\,dx + v_y\,dy) + i\oint (v_x\,dy - v_y\,dx). \tag{52}$$

The real part represents the circulation round the boundary; for the irrotational flow under consideration this should vanish. The imaginary part represents the total rate of flow across the boundary, and must equal the sum of emissions from simple sources inside the boundary, that is, the sum of the residues multiplied by 2π. A multiple source in the sense just defined contributes nothing to the integral since the liquid emitted by the simple sources constituting the multiple source is returned to the sink.

Laurent's series is a representation of any given complex velocity by an equivalent combination of sources of various orders situated at $z = a$ and $z = \infty$. The sources at infinity are represented by the non-negative powers of $z - a$.

A similar situation exists in the case of a solenoidal flow, as in a magnetic field produced by parallel electric current filaments. Around a single filament the magnetic intensity is

$$H_\rho = 0, \qquad H_\varphi = I/2\pi\rho, \tag{53}$$

where I is the current. The circulation of H round the circle of radius ρ, coaxial with the filament, is I; the outflow of H across the circle is zero. In this case

$$H_x - iH_y = I/2\pi iz. \tag{54}$$

If the source distributions in the irrotational and solenoidal fields are the same, the fields are alike except for the interchange of equipotential and stream lines; either type of source illustrates the nature of the poles of functions of a complex variable.

9. *Applications of the theorem of residues to the calculation of integrals of functions of a real variable*

Sometimes it is advantageous to reduce integrals of functions of a real variable to integrals in the complex plane in order to utilize the simplifications offered by the theorem of residues.

Consider the following integral

$$P = \int_0^{2\pi} \frac{d\varphi}{1 + k\cos\varphi}, \qquad -1 < k < 1. \tag{55}$$

Since $\cos\varphi = \frac{1}{2}(z + z^{-1})$, where z is on the unit circle, we have ($d\varphi = dz/iz$)

$$P = -2ik^{-1} \oint \frac{dz}{z^2 + 2k^{-1}z + 1}\cdot \tag{56}$$

The poles of the integrand are at

$$z_1 = -k^{-1} + \sqrt{k^{-2} - 1}, \qquad z_2 = -k^{-1} - \sqrt{k^{-2} - 1}. \tag{57}$$

The product z_1z_2 is unity. Since both quantities, z_1 and z_2, are real, only one pole is inside the unit circle. If k is positive, z_1 is inside the circle and the residue is

$$\begin{aligned} \operatorname{Res} 1/(z^2 + 2k^{-1}z + 1)_{z=z_1} &= \lim \frac{z - z_1}{(z - z_1)(z - z_2)} = \frac{1}{z_1 - z_2} \\ &= \frac{1}{2\sqrt{k^{-2} - 1}}\cdot \end{aligned} \tag{58}$$

Therefore,

$$P = 2\pi/\sqrt{1 - k^2}. \tag{59}$$

If k is negative, z_2 is inside the unit circle; but the answer is the same.

When k is greater than unity, the integral diverges. In this case z_1 and z_2 are complex and *on* the circle; and the theorem of residues is not applicable.

For the next example take the Fresnel integrals

$$\begin{aligned} P_1 &= \int_0^\infty \cos x^2\, dx = \tfrac{1}{2}\int_0^\infty \frac{\cos x}{\sqrt{x}}\, dx, \\ P_2 &= \int_0^\infty \sin x^2\, dx = \tfrac{1}{2}\int_0^\infty \frac{\sin x}{\sqrt{x}}\, dx. \end{aligned} \tag{60}$$

Form a complex integral

$$2P = 2P_1 + 2iP_2 = \int_0^\infty \frac{e^{ix}}{\sqrt{x}}\, dx. \tag{61}$$

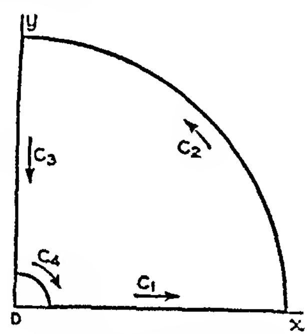

FIG. 15.7. A contour of integration for the integral (61).

If C_1 is the positive real axis, Figure 15.7, this integral may be expressed as

$$2P = \int_{C_1} \frac{e^{iz}}{\sqrt{z}}\, dz. \tag{62}$$

Anywhere inside the first quadrant indented at the origin O, the integrand

is finite and single-valued; therefore

$$\int_{C_1} + \int_{C_2} + \int_{C_3} + \int_{C_4} = 0. \tag{63}$$

As the radius r of the indentation C_4 approaches zero, the fourth integral approaches zero as the square root of r. As the radius R of the circular part C_2 increases, the corresponding integral diminishes; for the integrand diminishes as exp $(-y)$. Therefore,

$$2P = -\int_{C_3} \frac{e^{iz}}{\sqrt{z}}\,dz = i\int_0^\infty \frac{e^{-y}}{\sqrt{iy}}\,dy. \tag{64}$$

Starting with positive real values of $\sqrt{z}$ on C_1, we arrive, via C_2, at the value $\sqrt{y}$ exp $(i\pi/4)$ on C_3; that is, $\sqrt{i} = (1+i)/\sqrt{2}$ and not the negative of it. Substituting $y = t^2$, and referring to (6–45), we obtain

$$2P = \sqrt{2}(1+i)\int_0^\infty e^{-t^2}\,dt = \sqrt{\pi/2}(1+i). \tag{65}$$

Comparing with (61), we have

$$P_1 = P_2 = \sqrt{\pi/8}. \tag{66}$$

In retrospect, we find that the present method depends on deforming the contour of integration without changing the value of the integral. In terms of the real part of z, the integrand assumes different forms, depending on the contour of integration. Thus we obtain several integrals which are equal to each other, and the value of one may happen to be known. Even if all the integrals are unknown, we may find that some can be evaluated more easily than others. When we are forced to resort to numerical integration, we may find that some of these integrals can be evaluated more rapidly than others. The integrands in (60) assume alternately positive and negative values as x increases, and the convergence is slow; but the integrand in (65) decreases very rapidly and numerical integration would have to be extended over a relatively short interval.

Problems

1. Integrate $I = \int_0^\infty \frac{dx}{x^2 + a^2}$ with the aid of the theorem of residues. *Ans.* $\pi/2a$.

Hint: Note that $2I = \int_{-\infty}^{\infty} \frac{dx}{x^2+a^2}$; close the real axis with an infinite half-circle and prove that this half-circle contributes nothing to the value of the integral.

2. Evaluate $\int_0^\infty \frac{dx}{x^4 + a^4}$. *Ans.* $\pi\sqrt{2}/4a^3$.

3. Evaluate $\int_0^\infty \frac{\cos x\,dx}{x^2 + a^2}$. *Ans.* $\pi e^{-a}/2a$.

Hint: Integrate $e^{iz}/(z^2 + a^2)$ round the contour used in the preceding problems.

4. Evaluate $\int_0^\infty \frac{dx}{x^{2n} + a^{2n}}$, where n is a positive integer.

Ans. $\pi/2na^{2n-1} \sin(\pi/2n)$.

5. Consider $\int_0^\infty e^{-z^2}\,dz$. Show that the contour of integration (the positive real axis) can be rotated, without changing the value of the integral, through an angle α not exceeding 45°. Hence obtain

$$\int_0^\infty \exp(-x^2\cos 2\alpha)\cos(x^2\sin 2\alpha)\,dx = \tfrac{1}{2}\sqrt{\pi}\cos\alpha,$$

$$\int_0^\infty \exp(-x^2\cos 2\alpha)\sin(x^2\sin 2\alpha)\,dx = \tfrac{1}{2}\sqrt{\pi}\sin\alpha.$$

6. Show that $\oint z^{-n-1}e^z dz = 2\pi i/n!$ for any contour enclosing $z = 0$ ($n = 0$, $1, 2, 3\cdots$). If the contour is chosen to be a circle of radius a, then

$$\int_0^{2\pi} e^{a\cos\varphi}\cos(a\sin\varphi - n\varphi)\,d\varphi = 2\pi a^n/n!$$

$$\int_0^{2\pi} e^{a\cos\varphi}\sin(a\sin\varphi - n\varphi)\,d\varphi = 0.$$

7. Integrating e^{-z}/z round the contour in Figure 15.7, show that

$$\int_0^\infty \frac{\sin y}{y}\,dy = \tfrac{1}{2}\pi.$$

10. *Integrals of multiple-valued functions*

When integrating single-valued functions, we have to think of poles only. If the contour of integration is the boundary of a simply-connected region, two cases present themselves:

(a) There are no poles in the region; then the integral is zero.
(b) There are poles in the region; then the integral in the counter-clockwise direction equals $2\pi i$ times the sum of the residues.

A multiply-connected region can always be converted into a simply-connected region by a proper number of cuts. In following the new contour of integration each cut is traversed twice, in opposite directions; and for *single-valued* functions the contributions from cuts are *always* equal to

zero. Because of this any closed contour of integration may be deformed without altering the value of the integral, so long as we do not slip over a pole. If we do, we should add or subtract the product of $2\pi i$ and the residue, depending on whether the pole is excluded from the region bounded by the contour or included in it — assuming, of course, that we are moving round the boundary in the counterclockwise direction. For the clockwise direction the " add " and " subtract " should be interchanged in this rule.

For an open curve the theorem of residues does not yield the value of the integral; it merely relates it to the integrals along other curves between the same end points.

In the case of multiple-valued functions the situation is complicated by the branch points. The theorem is applicable only to the boundaries of those simply-connected regions which do not contain any branch points. In general, this restriction makes it impossible to evaluate the integrals of multiple-valued functions solely in terms of the residues. However, the theorem of residues continues to be useful in simplifying the integrals.

Consider, for example,

$$P = \int \sqrt{1 - z^2}\, dz. \tag{67}$$

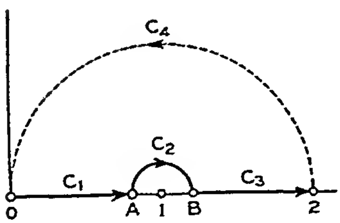

FIG. 15.8. A contour of integration for the integral (67).

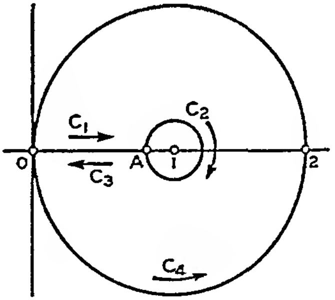

FIG. 15.9. An alternative contour for the integral (67).

The integrand has two branch points, $z = \pm 1$. On and inside the contours $C = C_1 + C_2 + C_3 + C_4$ in Figures 15.8 and 15.9, the integrand is single-valued and, there being no poles, the integral is zero. To prove the single-valuedness, let us examine the changes in the integrand along C. The integrand equals the square root of the product $(1 - z)(1 + z)$ and its phase is half the sum of the phases of its factors. The phase of the integrand is all that we have to worry about; for, in this case, after the completion of the cycle the absolute value returns automatically to its

original value. The quantity $z_1 - z_2$ is represented by a vector joining z_2 to z_1; hence the phase of $1 + z$ is the inclination of the vector joining $z = -1$ to $z = z$. This phase remains constant as we move along C_1; it increases a little and then decreases to its original value as we go round C_2 in Figure 15.8; it remains constant on C_3; it varies on C_4 but returns to the original value at O. To summarize: the vector from $z = -1$ to $z = z$ swings as we go round C but it goes through the same total angle in the clockwise and counterclockwise directions and the net change in the phase is zero. In the case of $1 - z$, which is represented by the vector from $z = z$ to $z = 1$, the overall situation is the same; but there are some differences in detail. Along C_2 there is a net clockwise change in the phase which is equal to 180°; the phase of the square root changes by 90°; but in going round C_4 these changes are wiped out. Thus, the integral (67) along $C = C_1 + C_2 + C_3 + C_4$ vanishes; and the integral along $C_1 + C_2 + C_3$ is equal to that along $-C_4$.

If we wish to evaluate (67) along an open contour $C_1 + C_2 + C_3$, we note that the radius of the half-circle C_2 can be made infinitely small without affecting the value of the integral. This is because in deforming the contour we shall not pass through any singularities. However, we cannot dispense with the infinitely small indentation even though the integral round it turns out to be zero. If we were to pass directly through $z = 1$, we should be in a quandary about the sign of the square root; the knowledge of the sign at O would not help us and we should have to assign the sign arbitrarily; that is, in effect, the integrals along C_1 and C_3 would become independent. The indentation, on the other hand, connects the two integrals unambiguously. Suppose we start with $\sqrt{1 - z^2} = +1$ at $z = 0$; therefore, on C_1, $\sqrt{1 - z^2} = \sqrt{1 - x^2}$, where x is real and the square root is positive. We have already seen that in going round C_2 the phase diminishes by 90°; hence on C_3, $\sqrt{1 - z^2} = -i\sqrt{x^2 - 1}$, where x is real and the square root is positive. Therefore, along $C_1 + C_2 + C_3$ we have

$$P = \int_0^{1-r} \sqrt{1 - x^2}\, dx + \int_{C_2} \sqrt{1 - z^2}\, dz - i \int_{1+r}^{2} \sqrt{x^2 - 1}\, dx, \quad (68)$$

where r is the infinitely small radius of the indentation. On C_2 the absolute value of $1 - z$ is r; that of its square root is $\sqrt{r}$; the length of the contour is πr; and the absolute value of the integral is not greater than $\pi r\sqrt{r}$. Hence the integral approaches zero with r. Evaluating the other two integrals, we have

$$P = \tfrac{1}{4}\pi - i(\sqrt{3} - \tfrac{1}{2}\cosh^{-1} 2). \quad (69)$$

For the integral round the contour $C_1 + C_2 + C_3$ in Figure 15.9 the phase of $1 - z^2$ decreases by 360° in going round C_2; hence, the phase of the square root decreases by 180° and the integrand on C_3 is of opposite sign to that on C_1. Thus

$$\int_{C_1+C_2+C_3} \sqrt{1-z^2}\,dz = \int_0^1 \sqrt{1-x^2}\,dx - \int_1^0 \sqrt{1-x^2}\,dx$$
$$= 2\int_0^1 \sqrt{1-x^2}\,dx = \tfrac{1}{2}\pi. \tag{70}$$

If the initial value at O is taken as -1, the values of the above integrals will have opposite signs.

Next let us consider a multiple-valued integrand with a pole

$$Q = \int_C \frac{\sqrt{1-z^2}}{z-1.5}\,dz, \tag{71}$$

where C is the contour in Figure 15.9. The integrand is single-valued inside C and the theorem of residues is applicable. In evaluating the residue at $z = 1.5$ we must assign the proper value to $\sqrt{1-z^2}$. We have already found that, if we start with $+1$ at O, on the real axis for $x > 1$, $\sqrt{1-z^2} = -i\sqrt{x^2-1}$; therefore at $z = 1.5$, $\sqrt{1-z^2} = -i\sqrt{1.25} = -i\sqrt{5}/2$. Hence,

$$Q = \pi\sqrt{5}. \tag{72}$$

On the other hand, for the closed contour C_4,

$$R = \int_{C_4} \frac{\sqrt{1-z^2}}{z-1.5}\,dz, \tag{73}$$

the theorem of residues *is not applicable*, since there is a branch point inside C_4. Starting with a certain value of the integrand at O, we shall arrive at a different value on our return. It is important to understand the difference between this case and the preceding. There, after going round $C_1 + C_2 + C_3$, we also reach a different value of the integrand; but the entire path has not yet been completed. When it *is* completed via C_4, we arrive at the original value. Splitting C in (71) and inserting from (72) and (73), we obtain

$$\int_{C_1+C_2+C_3} + R = \pi\sqrt{5}, \qquad R = \pi\sqrt{5} - \int_{C_1+C_2+C_3} \tag{74}$$

The integral R has not been evaluated but merely reduced to another integral, which, depending on the circumstances, may or may not be easier to handle.

Problems

1. Calculate $\int \sqrt{z}\,dz$ along the real axis, from $z = -1$ to $z = 1$, indented over the origin. Let the initial value of the integrand be i. *Ans.* $\frac{2}{3}(1 + i)$.

Hint: For convenience introduce a new variable, $t = -z$, in the interval $(-1, -0)$; then $\sqrt{z} = i\sqrt{t}$, where $\sqrt{t}$ is positive real. If the initial value is assumed to be $-i$, then we should have $\sqrt{z} = -i\sqrt{t}$, where $\sqrt{t}$ is again positive real.

2. Solve the preceding problem if the contour is indented below the origin. *Ans.* $\frac{2}{3}(-1 + i)$.

3. Obtain the two preceding integrals from the antiderivative.

4. Evaluate $\displaystyle\int \frac{dz}{\sqrt{z}}$ and $\displaystyle\int \frac{dz}{z}$ along the contours suggested in the first two problems. In the first integral assume i as the initial value of the integrand. *Ans.* $2(i - 1)$, $2(1 + i)$; $-i\pi$, $i\pi$.

5. Evaluate $\int (1 - z^2)^{-1/2}\,dz$ along the real axis from $z = 0$ to $z = 2$, first indented over $z = 1$ and then under. Let unity be the initial value of the integrand. *Ans.* $\frac{1}{2}\pi + i\cosh^{-1} 2$, $\frac{1}{2}\pi - i\cosh^{-1} 2$.

6. Evaluate $\int z^{-1}(1 - z)^{1/2}\,dz$. The contour of integration runs along the lower side of the real axis from $z = -1$ to $z = -0$, around $z = 0$, and back to $z = -1$ along the upper side of the real axis. Let the initial value of the integrand be $\sqrt{2}$. *Ans.* $-2\pi i$.

7. In the preceding problem let the contour run from $z = -1$ to $z = 1 + 0$ along the lower side of the real axis and back along the upper side, with indentations around the pole and branch point. *Ans.* $4\sqrt{2} + 4\log(\sqrt{2} - 1)$.

8. Evaluate $\displaystyle\int \frac{dz}{\sqrt{4 - z} + \sqrt{1 + z}}$. The contour of integration runs along the lower real axis from $z = 0$ to $z = 4 + 0$ and back along the upper real axis, with indentations around $z = 1.5$ and $z = 4$. Let the initial value of the integrand be $\frac{1}{3}$.

Ans. $4 + \frac{1}{2}\sqrt{10}\log\dfrac{13 - 4\sqrt{10}}{3} + \frac{1}{2}i\pi\sqrt{10}$.

References

1. Thomas M. MacRobert, *Functions of a Complex Variable*, Macmillan and Company, London, 1925.

2. R. Rothe, F. Ollendorff and K. Pohlhausen (translation by Alfred Herzenberg), *Theory of Functions as Applied to Engineering Problems*, Technology Press, Massachusetts Institute of Technology, Cambridge, 1942.

CHAPTER XVI

Linear Analysis

1. *Linear systems*

A physical system is *linear* if its response to several causes acting in combination equals the sum of its responses to the same causes acting separately; that is, if the " principle of superposition " is applicable the system is linear. Mathematically, linear systems are described by linear differential and integral equations. Many important methods of solution of linear problems depend on the principle of superposition, do not apply to nonlinear problems, and thus may conveniently be collected under the heading " Linear Analysis." Some of these methods have been explained and used in earlier chapters; here we shall consider them from a broader point of view.

Causality is a physical concept and has no place in mathematics. As soon as physical relationships are translated into equations, there is only one difference between causes and responses: the former are supposed to be known functions and the latter unknown. The term " cause " becomes synonymous with the *given* or *independent function;* likewise the term " response " is applied to the *derived* or *dependent function.* In a sense, when we solve an equation we transform the given function into a new function which may be called the *transform* of the original function. Obviously, there is no essential difference between dependent and independent functions; either may be regarded as a transform of the other.

The function $y(t)$ is a *linear transform* of $x(t)$,

$$y(t) = \mathcal{L}[x(t)], \tag{1}$$

if

$$\mathcal{L}[x_1(t) + x_2(t)] = \mathcal{L}[x_1(t)] + \mathcal{L}[x_2(t)]. \tag{2}$$

If $0(t)$ is the " null function," that is, if $0(t)$ vanishes for all values of t,

$$\mathcal{L}[0(t) + 0(t)] = \mathcal{L}[0(t)] + \mathcal{L}[0(t)]. \tag{3}$$

Since $0(t) + 0(t) = 0(t)$, we have

$$\mathcal{L}[0(t)] = 0. \tag{4}$$

Thus, any linear transform of the null function is the null function: " there is no response without a cause."

Whenever equation (2) is satisfied, the operations performed on $x(t)$ are called linear. Thus multiplication by a constant is a linear operation

$$2[x_1(t) + x_2(t)] = 2x_1(t) + 2x_2(t); \tag{5}$$

that is $y(t) = 2x(t)$ is a linear transform of $x(t)$. The operation of differentiation, $\mathfrak{L} = d/dt$, is also a linear operation since

$$\frac{d}{dt}[x_1(t) + x_2(t)] = \frac{d}{dt}x_1(t) + \frac{d}{dt}x_2(t). \tag{6}$$

The integral

$$y(t) = \int_a^t x(\tau)\, d\tau, \tag{7}$$

where a is a constant, is a linear transform of $x(t)$ because

$$\int_a^t [x_1(\tau) + x_2(\tau)]\, d\tau = \int_a^t x_1(\tau)\, d\tau + \int_a^t x_2(\tau)\, d\tau. \tag{8}$$

Multiplication by a fixed function,

$$y(t) = a(t)x(t), \tag{9}$$

is a linear operation since

$$a(t)[x_1(t) + x_2(t)] = a(t)x_1(t) + a(t)x_2(t). \tag{10}$$

Similarly,

$$y(t) = \int_a^t K(t,\tau)x(\tau)\, d\tau, \tag{11}$$

where the *kernel* $K(t,\tau)$ is fixed, is a linear transform of $x(t)$ because

$$\begin{aligned}\int_a^t K(t,\tau)[x_1(\tau) + x_2(\tau)]\, d\tau &= \int_a^t K(t,\tau)x_1(\tau)\, d\tau \\ &+ \int_a^t K(t,\tau)x_2(\tau)\, d\tau.\end{aligned} \tag{12}$$

On the other hand, the operation of squaring is not a linear operation because

$$[x_1(t) + x_2(t)]^2 \neq [x_1(t)]^2 + [x_2(t)]^2. \tag{13}$$

The reader will recall that linear differential equations are defined as equations of the first degree in the dependent variable and its derivatives; thus

$$\frac{d^2y}{dt^2} + t^3\frac{dy}{dt} + e^ty = f(t) \tag{14}$$

is called a linear equation even though t^3, e^t and $f(t)$ are not linear functions. The principle of superposition is the key to this definition. In any linear nonhomogeneous differential equation the term not involving the dependent variable is a linear transform of the unknown function; thus in (14)

$$f(t) = \mathcal{L}[y(t)]. \tag{15}$$

When we try to recover $y(t)$ from $f(t)$ — that is, when we try to "solve" the equation — we find that the answer is not unique and that in general $y(t)$ is not a linear transform of $f(t)$; but we can subject $y(t)$ to supplementary conditions (initial or boundary conditions) which make it unique and a linear transform of $f(t)$. Then we could write the solution of (15) as the inverse transform

$$y(t) = \mathcal{L}^{-1}[f(t)]. \tag{16}$$

The supplementary conditions must be so fixed that the inverse transform of the null function is the null function, and that the solution of the homogeneous differential equation is identically equal to zero. The supplementary conditions have to be suitably chosen for each particular class of physical problem.

Let us consider a simple example. If a mass M is moving in a straight line under the influence of a force $F(t)$,

$$M\frac{dv}{dt} = F(t), \tag{17}$$

where $v(t)$ is the velocity. If $v(t) = v_1(t) + v_2(t)$, then clearly $F(t) = F_1(t) + F_2(t)$; and if the velocity is zero, the force acting on M is also zero. However, if v is constant, F is still equal to zero; and $v(t)$ can be recovered from (17) only except for an arbitrary constant. On further consideration we decide that if no force had ever acted on the body, the body would have remained at rest; that in computing the velocity resulting from the action of $F(t)$ we should use not only (17) but some supplementary condition. This condition is: *if $F(t)$ is finite and equal to zero when $t < \tau$, then $v(t)$ is continuous and equal to zero for $t < \tau$.* Thus we have summarized our physical ideas about the behavior of the body: the body is at rest if it has never been acted upon by a force; the velocity does not suddenly become different from zero if the force acting on the body is finite. Under these conditions the solution of (17) is unique

$$v(t) = \frac{1}{M}\int_\tau^t F(\xi)\, d\xi. \tag{18}$$

Let $F(t)$ be equal to zero except in the interval $(\tau, \tau + T)$ when $F(t) =$

p/T; then, from (18),

$$\begin{aligned} v(t) &= 0, && t \leq \tau; \\ &= \frac{p(t-\tau)}{MT}, && \tau \leq t \leq \tau + T; \\ &= \frac{p}{M}, && \tau + T < t. \end{aligned} \tag{19}$$

If $T \to 0$ and p is finite,

$$\begin{aligned} v(t) &= 0, && t \leq \tau; \\ &= p/M, && t > \tau. \end{aligned} \tag{20}$$

The velocity becomes discontinuous; but it has been produced by an infinite force. The quantity p is called *the impulse of force;* the product Mv is *the momentum;* and the two are equal.

If a mass M is attached to a spring of stiffness S and if s denotes the position of M with respect to the position in which the spring does not exercise any force on M,

$$M\frac{d^2s}{dt^2} + R\frac{ds}{dt} + Ss = F(t), \tag{21}$$

where the middle term represents the force due to friction. The supplementary conditions which make $s(t)$ a unique response to $F(t)$ and a linear transform of $F(t)$ are: *if $F(t) = 0$ for $t < \tau$ and is finite, then $s(t)$ and its first derivative are continuous and** $s(-\infty) = 0$. This makes $s(t) = 0$ the only solution of (21) when $F(t) = 0$.

More generally, if the dynamical system is described by an equation of order n, then the supplementary conditions consist in the continuity of all derivatives except the highest and the vanishing of the solution at $t = -\infty$. The highest derivative is continuous only if $F(t)$ starts from zero.

The nature of the supplementary conditions depends on the physical problem. Consider

$$\frac{d^2y}{dx^2} = y + f(x), \tag{22}$$

where y is the displacement in a one-dimensional wave. The general solution is

$$y(x) = Ae^x + Be^{-x} + g(x), \tag{23}$$

* In conjunction with the other supplementary conditions, the condition $s(t) = 0$ if $t < \tau$ is equivalent to $s(-\infty) = 0$.

where A and B are arbitrary constants and $g(x)$ is a particular solution depending on $f(x)$. Here we demand that $y(x)$ and $y'(x)$ be continuous and $y(\pm \infty) = 0$.

The electric potential V of a distribution of electric charge of density q satisfies the differential equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{q}{\epsilon}, \tag{24}$$

where ϵ is the dielectric constant. Physically V is uniquely defined by q; but (24) does not define V uniquely. The equation possesses numerous solutions even if $q = 0$. But if we specify that V and its partial derivatives are to be continuous when q is finite and that V vanishes at infinity, we shall find that V is uniquely determined by q and that it vanishes with it as it should.

The supplementary, initial or boundary, conditions are needed because we are usually unable to include all physical requirements in our differential equations.

Having supplemented our equations with the conditions sufficient to insure the unique solution of a given physical problem, we proceed as follows:

(a) We choose a set of simple independent functions in terms of which we can express any other given function;
(b) we obtain the solutions of our equations for this set of independent functions;
(c) and, finally, we obtain the solution for any given independent function by the principle of superposition.

There are three convenient basic sets of independent functions: a set of "unit step functions," a set of "unit impulse functions," and a set of sinusoidal or exponential functions. A given function may be expressed either as a succession of "steps" of infinitesimal or finite magnitude, or as a succession of "impulses," or as a "Fourier integral" of sinusoidal functions of variable frequency, or as a "Laplace integral" of exponential functions with a variable exponent. Laplace integrals are taken along certain contours in the complex plane and may often be evaluated by the calculus of residues.

These three sets of basic functions and their applications are considered successively in Sections 2, 3, 4. Any one of the three methods may prove best for any particular problem. As we shall see, we can express the responses to unit step functions and unit impulse functions by relatively simple Laplace integrals; but in simpler problems it is usually easier to find these responses directly. The Laplace transform method shows itself

to better advantage in more difficult problems. This is true of the other methods as well; simple problems are usually easier to solve by the more elementary methods.

2. The unit step function and the indicial admittance

The *unit step function*, Figure 16.1, is defined analytically as follows:

$$\begin{aligned} 1(t) &= 0, \quad t < 0; \qquad & 1(x) &= 0, \quad x < 0; \\ &= 1, \quad t > 0; & &= 1, \quad x > 0. \end{aligned} \tag{25}$$

Any piece-wise differentiable function $f(t)$ may be regarded as a succession of steps of finite or infinitesimal magnitude, $f'(t)\,dt$, Figure 16.2.

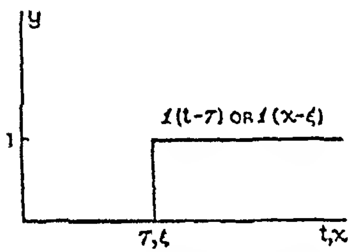

FIG. 16.1. The *unit step function*.

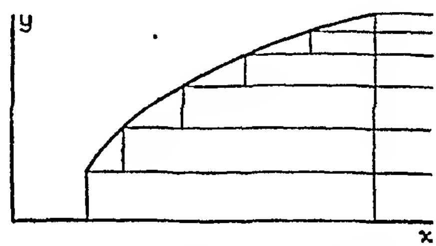

FIG. 16.2. The formation of an arbitrary function as a succession of unit step functions.

The response of a linear system to the unit step function is called the *indicial admittance $A(t,\tau)$ of the system.* In the case of time functions the indicial admittance vanishes if $t < \tau$ (no retroactive action),

$$A(t,\tau) = 0, \quad t < \tau. \tag{26}$$

We have seen that the coefficients of differential equations may be functions of the independent variable, t, and still be linear. The systems described by such equations are called *variable* linear systems to distinguish them from *invariable* systems, which are those governed by equations with constant coefficients. The indicial admittance of invariable systems depends only on $t - \tau$:

$$A(t,\tau) = A(t-\tau), \quad A(\bar{t}) = 0 \text{ if } \bar{t} < 0. \tag{27}$$

We are now in a position to write the response $F(t)$ to any piece-wise differentiable function $f(t)$ which vanishes prior to some particular instant,

$$F(t) = \int_{-\infty}^{t} A(t-\tau)\,df(\tau) = \int_{-\infty}^{\infty} A(t-\tau)\,df(\tau). \tag{28}$$

This integral is to be interpreted in the Stieltjes sense instead of the ordinary Riemann sense; that is, the differential is permitted to have finite values as well as infinitesimal values. If there are no discontinuities in $f(t)$ in the interval $(-\infty, t)$, (28) becomes an ordinary Riemann integral,

$$F(t) = \int_{-\infty}^{t} A(t - \tau)f'(\tau)\,d\tau. \tag{29}$$

If at some instant $t = t_n$, $f(t)$ changes suddenly from $f(t_n - 0)$ to $f(t_n + 0)$, (28) becomes

$$F(t) = \sum_n [f(t_n + 0) - f(t_n - 0)]A(t - t_n) + \int_{-\infty}^{t} A(t - \tau)f'(\tau)\,d\tau, \tag{30}$$

where the integral is now interpreted in the Riemann sense and the summation is extended over all points of discontinuity.

The general problem is thus reduced to the calculation of the indicial admittance of the system. To illustrate, let us consider an electric circuit consisting of an inductance L in series with a resistance R. To obtain the indicial admittance, that is, the current in response to the unit step voltage, we have to solve

$$L\frac{dA}{dt} + RA = 1(t). \tag{31}$$

For $t > 0$, this equation is simply

$$L\frac{dA}{dt} + RA = 1. \tag{32}$$

The general solution is the sum of a particular solution and the complementary function (that is, the general solution of the associated homogeneous equation). If $A =$ constant, $dA/dt = 0$, and $A = 1/R$ is seen to be a particular solution; thus

$$A(t) = C\exp(-Rt/L) + 1/R. \tag{33}$$

The constant of integration C is found from the condition that the current vanishes at $t = 0$; hence

$$C + 1/R = 0, \qquad C = -1/R,$$
$$A(t) = (1/R)[1 - \exp(-Rt/L)]. \tag{34}$$

There might be some doubt as to the vanishing of the current at $t = 0$. The current due to the unit step voltage must certainly be zero before the voltage begins to act, that is, when $t < 0$; and if the current is continuous, of course it vanishes at $t = 0$. However, can we safely assume that the

current is continuous? Let us go back to (31) and integrate it from $t = -0$ to $t = t$:

$$L[A(t) - A(-0)] + R\int_{-0}^{t} A\,dt = \int_{-0}^{t} 1(t)\,dt = t. \tag{35}$$

As $t \to +0$, the right side approaches zero; so does the second term on the left. Hence $A(+0) = A(-0)$ as long as $L \neq 0$. If $L = 0$ the argument breaks down, and there seems to be nothing wrong with an assumption of a sudden rise in current; but then the problem is altered and there is no arbitrary constant of integration to be determined.

The response of the circuit to an arbitrary voltage $V(t)$ is then written down from equation (28),

$$I(t) = \frac{1}{R}\int_0^t \{1 - \exp[-R(t-\tau)/L]\}\,dV(\tau). \tag{36}$$

If $V(t)$ is a square pulse of unit magnitude, beginning at $t = 0$ and ending with $t = T$, then, within the interval $0 < t < T$, $dV(\tau) = 0$ and the contributions come only from the discontinuities

$$\begin{aligned} I(t) &= 0, \qquad t < 0; \\ &= \frac{1}{R}\{1 - \exp[-Rt/L]\}, \qquad 0 \leq t \leq T; \\ &= \frac{1}{R}\{\exp[-R(t-T)/L] - \exp[-Rt/L]\}, \qquad t \geq T. \end{aligned} \tag{37}$$

If $V(t) = \sin(\pi t/T)$ when $0 \leq t \leq T$ and vanishes outside the interval,

$$I(t) = \frac{\pi}{TR}\int_0^t \{1 - \exp[-R(t-\tau)/L]\}\cos(\pi\tau/T)\,d\tau. \tag{38}$$

Only routine integration is needed to evaluate $I(t)$.

Let us now consider the case of a capacitance C in series with a resistance R. The equation for the indicial admittance is

$$RA + C^{-1}\int_0^t A\,dt = 1(t). \tag{39}$$

The lower limit of the integral is taken to be zero in order to insure that there is no charge on the capacitor at $t = 0$. When $t > 0$, the right side of this equation is unity; differentiating, we have

$$R\frac{dA}{dt} + \frac{A}{C} = 0. \tag{40}$$

Solving, we obtain

$$A(t) = P \exp(-t/RC), \tag{41}$$

where P is an arbitrary constant. This time the current and A are not continuous at $t = 0$; thus we find directly from (39) that $A(+0) = 1/R$, and

$$A(t) = (1/R) \exp(-t/RC). \tag{42}$$

If we are interested in the charge $q(t)$ on the capacitor instead of the current $I(t) = dq/dt$, the equation for the indicial admittance becomes

$$R\frac{dA}{dt} + \frac{A}{C} = 1(t). \tag{43}$$

This is essentially the same equation as (31) and the same solution will be valid, with obvious minor changes.

If we *introduce a new variable of integration*, $\hat{\tau} = t - \tau$, *in (30)* and then drop the caret superscript, we obtain

$$F(t) = \int_0^\infty A(\tau)f'(t-\tau)\,d\tau + \sum_n [f(t_n+0) - f(t_n-0)]A(t-t_n), \tag{44}$$

where the summation is extended over the set of points of discontinuity $t_1, t_2, t_3, \cdots$ of $f(t)$. This is a frequently used form of the general solution in terms of the indicial admittance.

Problems

1. Find the indicial admittance of a circuit consisting of an inductance L, resistance R, and capacitance C in series; that is, solve $L\dfrac{dA}{dt} + RA + \dfrac{1}{C}\displaystyle\int_0^t A\,dt = 1(t)$.

Hint: To determine the arbitrary constants prove that A vanishes at $t = 0$ and its derivative suddenly rises by an amount $1/L$.

Ans. $A(t) = (e^{p_1 t} - e^{p_2 t})/(p_1 - p_2)L$, where p_1 and p_2 are the zeros of $LCp^2 + RCp + 1$.

2. Find the indicial admittance of the circuit in Problem 1 if the charge on the condenser is regarded as the dependent function.

Ans. $A(t) = [C/(p_1 - p_2)](p_1 - p_2 + p_2 e^{p_1 t} - p_1 e^{p_2 t})$.

3. Find the indicial admittance of a circuit consisting of a resistance R in series with a parallel combination of an inductance L and capacitance C.

Note: It is very likely that the reader will run into difficulties and the problem is a challenge to him. If he fails to resolve the difficulties, he is urged to try again after he learns the answer, given in a later section, where this problem is solved by the Laplace transform method.

4. Find the response (that is, the voltage) of a circuit consisting of L, R, C in parallel to a unit step current. Note the similarity between this problem and Problem 1.

3. *The unit impulse function and Green's function*

The *unit impulse function* and the generation of arbitrary functions by a succession of impulse functions of varying moments are shown in Figures 16.3 and 16.4. The analytic definition of the unit impulse function is:

$$
\begin{aligned}
I(t,\tau) &= 0, \quad \text{if} \quad t < \tau \quad \text{or} \quad t > \tau; \\
&= \infty, \ \text{if} \quad t = \tau; \\
\int_{\tau-0}^{\tau+0} I(t,\tau)\, dt &= 1.
\end{aligned}
\tag{45}
$$

In a more extended form,

$$
I(t,\tau) = \lim \frac{1(t - \tau + \frac{1}{2}s) - 1(t - \tau - \frac{1}{2}s)}{s}, \qquad s \to 0; \tag{46}
$$

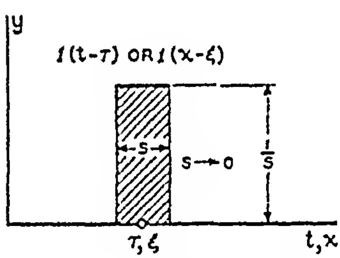

Fig. 16.3. The *unit impulse function.*

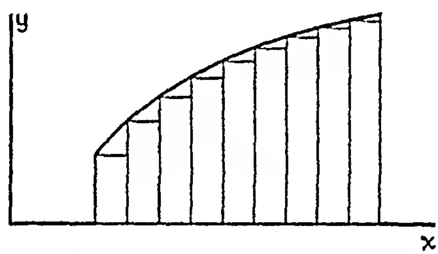

Fig. 16.4. The formation of an arbitrary function as a succession of unit impulse functions.

that is, the impulse function is the limit of two opposite step functions of indefinitely increasing magnitude such that *the moment of the impulse*, defined as the area under the step functions, is unity. For any function $f(t)$,

$$
\begin{aligned}
\int_{t_1}^{t_2} f(t) I(t,\tau)\, dt &= f(\tau), \quad t_1 < \tau < t_2; \\
&= 0, \qquad \tau < t_1 \quad \text{or} \quad \tau > t_2.
\end{aligned}
\tag{47}
$$

If τ takes on different values while t_1, t_2 are kept fixed, the integral represents a function which is identical with $f(t)$ inside the interval and vanishes outside it.

The response $G(t,\tau)$ of a system to the unit impulse function is called *Green's function*. For functions of time

$$G(t,\tau) = 0, \qquad t < \tau, \tag{48}$$

there being no retroactive action. For invariable systems

$$G(t,\tau) = G(t - \tau), \qquad G(t) = 0 \qquad \text{if } t < 0. \tag{49}$$

If the independent coordinate is spatial, Green's functions may be of varied forms since the boundary conditions are varied. For example, if a unit electromotive force is impressed at $x = \xi$ on a transmission line infinite in both directions, the current in the line is

$$\begin{aligned} I(x,\xi) = G(x,\xi) &= \tfrac{1}{2}K^{-1} \exp[-i\beta(x - \xi)], \qquad x > \xi; \\ &= \tfrac{1}{2}K^{-1} \exp[-i\beta(\xi - x)], \qquad x < \xi. \end{aligned} \tag{50}$$

The condition which results in this particular solution of the transmission line equations is that the phase of the current is *retarded* with increasing distance from the source. If the line is made slightly dissipative, the condition at infinity is easier to understand; all solutions except one become infinite there and thus do not correspond to our idea of a possible physical behavior of the system. If the dissipation is made smaller and smaller, the "physical" solution approaches (50).

The portion of the transmission line to the left of $x = 0$ may be cut off; then the response $G(0,\xi)$ must vanish at $x = 0$. This condition replaces the one which required G to vanish at $x = -\infty$ in the dissipative case, or, in the nondissipative case, to be a retarded wave function. The new Green's function is

$$\begin{aligned} G(x,\xi) &= iK^{-1} \sin \beta\xi \exp(-i\beta x), \qquad x > \xi; \\ &= iK^{-1} \sin \beta x \exp(-i\beta\xi), \qquad x < \xi. \end{aligned} \tag{51}$$

If a shunt generator of infinite impedance is supplying current K^{-1} amperes to an infinite line at $x = \xi$,

$$\begin{aligned} I(x,\xi) = G(x,\xi) &= \tfrac{1}{2}K^{-1} \exp[-i\beta(x - \xi)], \qquad x > \xi; \\ &= -\tfrac{1}{2}K^{-1} \exp[-i\beta(\xi - x)], \qquad x < \xi. \end{aligned} \tag{52}$$

The Green's function for the double source consisting of properly related series and shunt generators is obtained by adding (50) and (52). It vanishes if $x < \xi$ and in this respect resembles Green's functions for functions of time — that is, certain double sources act only to the right and not to the left.

In the later sections we shall deal with Green's functions of the type (50), (51) and (52) in more detail; but for the present let us return to problems with time t as the independent variable.

Once the Green's function for the given system of equations and boundary conditions has been determined, the response $F(t)$ to an arbitrary distribution of sources $f(t)$ is obtained by superposition

$$F(t) = \int_{-\infty}^{\infty} G(t,\tau)f(\tau)\,d\tau. \tag{53}$$

Aside from the evaluation of this integral, which is very difficult at times, the main problem is the calculation of Green's functions.

From the definition of the unit impulse function in terms of the unit step functions we obtain

$$G(t,\tau) = -\frac{\partial A(t,\tau)}{\partial \tau}. \tag{54}$$

The method of Green's functions is essentially the same as the method of indicial admittances and, in general, more convenient because the equations which must be solved in order to obtain the Green's function are homogeneous for all values of the independent variable.

The direct method of calculation of Green's functions has already been illustrated in the chapters on differential equations. We shall consider, however, one or two examples for the purposes of comparison with the indicial admittance method. For a series circuit consisting of resistance R and inductance L, we have

$$L\frac{dG}{dt} + RG = I(t,\tau). \tag{55}$$

This is similar to (31) in the indicial admittance method. Equation (55) is really

$$L\frac{dG}{dt} + RG = 0, \qquad t \neq \tau, \tag{56}$$

with a certain supplementary condition. Since the point $t = \tau$ is excluded in (56), we are forced to treat the intervals $(-\infty,\tau)$ and (τ,∞) separately, and write the general solution in the form

$$\begin{aligned} G &= P \exp(-Rt/L), \qquad t < \tau; \\ &= Q \exp(-Rt/L), \qquad t > \tau; \end{aligned} \tag{57}$$

containing two constants of integration instead of one. Since the circuit is not expected to respond before the voltage impulse is applied, $P = 0$. The constant Q is obtained from (45); the impulse function is not defined at $t = \tau$ but its integral is. Integrating (55) from $t = \tau - 0$ to $t = \tau + 0$, we have

$$LG(\tau + 0,\tau) - LG(\tau - 0,\tau) = 1. \tag{58}$$

Substituting from (57),

$$LQ \exp(-R\tau/L) - LP \exp(-R\tau/L) = 1,$$
$$LQ \exp(-R\tau/L) = 1, \qquad Q = L^{-1} \exp(R\tau/L); \tag{59}$$

thus

$$G(t,\tau) = 0, \qquad t < \tau,$$
$$= L^{-1} \exp[R(\tau - t)/L], \qquad t > \tau. \tag{60}$$

The second example of the preceding section concerns a series circuit consisting of resistance R and capacitance C. Instead of (39) we now have

$$RG + C^{-1} \int_\tau^t G(t)\, dt = I(t,\tau). \tag{61}$$

In this case Green's function does not exist, since it is only the integral of $I(t,\tau)$ that is defined at $t = \tau$, and if we integrate (61) we obtain

$$R\int_{\tau-0}^{\tau+0} G\, dt + C^{-1} \int_{\tau-0}^{\tau+0} dt \int_\tau^t G(t)\, dt = 1. \tag{62}$$

If G is finite at $t = \tau$, the integrals on the left vanish; hence G is infinite at $t = \tau$ and we cannot determine the arbitrary constants in the solution of the homogeneous equations for $t \neq \tau$.

The above result is not surprising if we note that for the impulsive voltage the initial current is $1/L$, where L is the inductance of the circuit. This conclusion can be drawn from (60) and it is not affected by the addition of capacitance. Two alternatives now present themselves: (1) to use the indicial admittance method, (2) to alter the dependent function from the current $I(t)$ to the charge $q(t)$ on the capacitor, given by $I = dq/dt$. Let us see how the second alternative works. The equation becomes

$$R\frac{d\bar{G}}{dt} + \frac{\bar{G}}{C} = I(t,\tau). \tag{63}$$

By integration we have, just as in the preceding example (with L and R),

$$\bar{G}(\tau + 0,\tau) - \bar{G}(\tau - 0,\tau) = 1/R. \tag{64}$$

This equation tells us that there is a sudden rise of charge on the condenser from zero to $1/R$. The solution itself can be written by analogy with (60):

$$\bar{G}(t,\tau) = R^{-1} \exp[(\tau - t)/RC], \qquad t > \tau. \tag{65}$$

For any voltage $V(t)$,

$$q(t) = R^{-1} \int_{-\infty}^{\infty} V(\tau) \exp[(\tau - t)/RC]\, d\tau, \qquad t > \tau. \tag{66}$$

The current is then obtained by differentiation,

$$I(t) = -R^{-2}C^{-1}\int_{-\infty}^{\infty} V(\tau)\exp[(\tau - t)/RC]\,d\tau, \qquad t > \tau. \tag{67}$$

Let us now return to the voltage impulse and assume that $V(\tau)$ increases while the action interval decreases so that the volt-seconds remain constant and equal to unity; then from (67)

$$I(t) = -R^{-2}C^{-1}\exp[(\tau - t)/RC], \qquad t > \tau. \tag{68}$$

This seems to be the response to the unit voltage impulse; but how is it that we could not obtain it directly from (61) and (62)? Is there any inconsistency? There is no inconsistency because (68) is not quite the same function as G. Equation (64) tells us that in passing through $t = \tau$, the instant of the impulse, there is a sudden shift of electric charge from one plate to the other. This implies an infinite current at $t = \tau$. Then the impulse is over and the discharge of the capacitor begins; during this discharge we have (68).

Let us now see what happens when the circuit contains inductance as well as resistance and capacitance. The equation of oscillations is

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int_{-\infty}^{t} I\,dt = V(t). \tag{69}$$

For a unit impulse at $t = 0$, the equation becomes

$$L\frac{dG}{dt} + RG + \frac{1}{C}\int_{-\infty}^{t} G\,dt = I(t,0). \tag{70}$$

Integrating over $-0 < t < +0$, we have

$$LG(+0,0) - LG(-0,0) = 1, \qquad G(+0,0) = 1/L. \tag{71}$$

Elsewhere $G(t,0)$ is a solution of the associated homogeneous equation,

$$\begin{aligned} G(t,0) &= Ae^{p_1 t} + Be^{p_2 t}, \qquad t > 0;\\ &= 0, \qquad t < 0; \\ LCp_{1,2}^2 &+ RCp_{1,2} + 1 = 0. \end{aligned} \tag{72}$$

From (71) and (72),

$$A + B = 1/L. \tag{73}$$

To obtain the second equation for the arbitrary constants, we calculate the charge on the capacitor,

$$\int_0^t G(t,0)\,dt = Ap_1^{-1}e^{p_1 t} - Ap_1^{-1} + Bp_2^{-1}e^{p_2 t} - Bp_2^{-1}. \tag{74}$$

At $t = \infty$ the capacitor should be discharged; since the exponential terms vanish, we have

$$Ap_1^{-1} + Bp_2^{-1} = 0, \qquad B = -(p_2/p_1)A. \tag{75}$$

Substituting in (73), solving for A and B, and inserting their values in (72), we obtain the Green's function

$$G(t,0) = \frac{p_1 e^{p_1 t} - p_2 e^{p_2 t}}{L(p_1 - p_2)}. \tag{76}$$

Hence the general response is

$$\begin{aligned} I(t) &= \frac{1}{L(p_1 - p_2)} \int_{-\infty}^{t} [p_1 e^{p_1(t-\tau)} - p_2 e^{p_2(t-\tau)}] V(\tau)\, d\tau \\ &= \frac{1}{L(p_1 - p_2)} \int_0^{\infty} \{p_1 e^{p_1 \tau} - p_2 e^{p_2 \tau}\} V(t - \tau)\, d\tau. \end{aligned} \tag{77}$$

A variant procedure is to express (69) in terms of the charge $q(t)$; the boundary conditions are then

$$G(+0,0) = 0, \qquad G'(+0,0) = 1/L, \tag{78}$$

where $G(t,0)$ is the response to the unit impulse at $t = 0$. This time we do not have to anticipate a part of the answer by deciding on physical grounds what should happen at $t = \infty$.

Differential equations of invariable systems are invariant under translation of the origin of time and $G(t,\tau) = G(t - \tau)$; hence (53) may be written,

$$\begin{aligned} F(t) &= \int_{-\infty}^{\infty} G(t - \tau) f(\tau)\, d\tau = \int_{-\infty}^{t} G(t - \tau) f(\tau)\, d\tau \\ &= \int_0^{\infty} G(\tau) f(t - \tau)\, d\tau. \end{aligned} \tag{79}$$

The upper limit in the first integral is replaced by t in the second since G vanishes when $\tau > t$.

Problems

1. Find the response of a circuit consisting of R and C in parallel to a unit impulse of current.

Ans. The voltage across the circuit is $V(t) = C^{-1} \exp(-t/RC)$.

2. Find the response of a circuit consisting of R and L in parallel to a unit impulse of current. *Hint:* Introduce the magnetic flux $\Phi = \int^t V(t)\, dt$ through the inductance as the dependent variable.

Ans. There is a momentary infinite current in the resistance and a corresponding momentary infinite voltage across the inductance. The latter establishes instantaneously a magnetic flux, R volt-seconds, through the inductance. Thereafter, the magnetic flux disappears in accordance with $\Phi = R \exp(-Rt/L)$, and the voltage across the inductance (in the direction of the current impulse) is

$$V = -(R^2/L) \exp(-Rt/L).$$

3. What is the current in the inductance in Problem 2 when $t > 0$? *Note:* The current *entering* the circuit is zero when $t > 0$ since the current impulse has terminated but the current *in* the circuit is not zero.
Ans. $I_L(t) = (R/L) \exp(-Rt/L)$.

4. Find the response of a circuit consisting of L, R, C in parallel to a unit current impulse.

4. *The exponential function and Laplace transforms*

In the more complicated cases the determination of either the indicial admittance or the Green's function becomes laborious. If the coefficients in the equations are variable there is no alternative; but when the coefficients are constant, "Laplace integrals" come to the rescue. The basic idea is to express all functions $f(t)$, the known and the unknown, as integrals of the form $\int \bar{f}(p) \exp(pt)\, dp$. When these *Laplace integrals* are substituted in the equations, the problem becomes considerably simplified.

First let us consider a method for expressing known functions as Laplace integrals. The knowledge of Cauchy's integral formula and a little imagination should suggest that the unit step function is given by the following contour integral

$$1(t-\tau) = \frac{1}{2\pi i}\int_C \frac{e^{p(t-\tau)}}{p}\, dp, \qquad (80)$$

where the contour of integration is the imaginary axis indented to the right of the origin, Figure 16.5, or any other contour obtained from this by continuous deformation, without crossing the origin and without leaving the imaginary axis at infinity. If $t < \tau$, $t - \tau$ is negative and, at infinity, the exponential function vanishes when the real part of p is positive — that is, everywhere on the infinite semicircle C_R in the right half-plane, Figure 16.5. It can be shown that the contribution to the

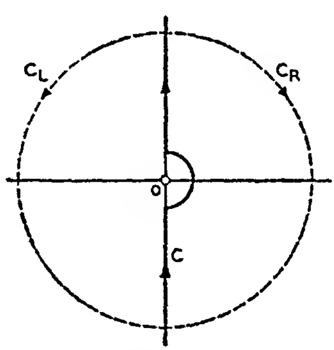

Fig. 16.5. The contour C of integration in the integral representation of the unit step function.

integral from C_R is zero and we can add C_R to C without affecting the value of the function represented by (80). Inside the closed contour thus formed, the integrand is analytic and the integral is zero.

If $t > \tau$, $t - \tau$ is positive and the exponential function vanishes on the infinite semicircle C_L in the left half-plane. The integrand has one pole, $p = 0$, inside $C + C_L$; hence, by the theorem of residues the value of the integral is unity.

Next we shall represent a function $f(t)$ which vanishes at $t = \mp\infty$ by an integral of the form

$$f(t) = \int_C S(p)e^{pt}\,dp. \tag{81}$$

The function $S(p)$ is called the *Laplace transform of* $f(t)$. Frequently $1/2\pi i$ is inserted as a factor in front of the integral in (81). The Laplace transform of the unit step function is

$$S(p) = e^{-p\tau}/2\pi ip. \tag{82}$$

When $1/2\pi i$ is exhibited in front of the integral, the Laplace transform of the unit step is $e^{-p\tau}/p$.

Regarding $f(t)$ as a succession of steps of magnitude $df(t)$, we have

$$f(t) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} df(\tau) \int_C \frac{e^{p(t-\tau)}}{p}\,dp. \tag{83}$$

Assuming that we can interchange the order of integration, we obtain

$$f(t) = \frac{1}{2\pi i}\int_C \frac{e^{pt}\,dp}{p} \int_{-\infty}^{\infty} e^{-p\tau}\,df(\tau). \tag{84}$$

Integrating by parts,

$$\begin{aligned} f(t) &= \frac{1}{2\pi i}\int_C \frac{e^{pt}}{p}\,dp\left[e^{-p\tau}f(\tau)\Big|_{-\infty}^{\infty} + p\int_{-\infty}^{\infty} f(\tau)e^{-p\tau}\,d\tau\right] \\ &= \frac{1}{2\pi i}\int_C e^{pt}\,dp\left[\int_{-\infty}^{\infty} f(\tau)e^{-p\tau}\,d\tau\right]. \end{aligned} \tag{85}$$

This equation is true whether $f(t)$ is continuous or not. If $f(t)$ has a discontinuity, a finite step must be added in the first line of (85); however, this term is canceled by a corresponding term when we evaluate $e^{-p\tau}f(\tau)$. Comparing (81) and (85) we have the Laplace transform of $f(t)$

$$S(p) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} f(\tau)e^{-p\tau}\,d\tau. \tag{86}$$

Since $f(\tau)$ is supposed to vanish prior to some instant τ, we can always

choose the origin of time to make the function equal to zero in the interval $(-\infty, 0)$; then

$$S(p) = \frac{1}{2\pi i}\int_0^\infty f(\tau)e^{-p\tau}\,d\tau. \tag{87}$$

Once the Laplace transform of the independent function has been found, the dependent function can also be expressed as a Laplace integral. Consider the following equation (with constant coefficients):

$$a_n\frac{d^ny}{dt^n} + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1\frac{dy}{dt} + a_0y = f(t). \tag{88}$$

First we solve this equation subject to the condition that $y(t)$ and $f(t)$ are exponential functions

$$y(t) = \mathfrak{Y}e^{pt}, \qquad f(t) = Ae^{pt}. \tag{89}$$

Substituting in (88), we find

$$\mathfrak{Y} = \frac{A}{Z(p)}, \qquad Z(p) = a_np^n + a_{n-1}p^{n-1} + \cdots + a_1p + a_0. \tag{90}$$

In electrical problems $Z(p)$ is referred to as the "impedance function." By superposition of responses to $A = S(p)\,dp$, we obtain

$$y(t) = \int_C \frac{S(p)}{Z(p)}e^{pt}\,dp. \tag{91}$$

In a more formal analysis we might start with Laplace integrals

$$f(t) = \int_C S(p)e^{pt}\,dp, \qquad y(t) = \int_C U(p)e^{pt}\,dp, \tag{91A}$$

in which $S(p)$ is supposed to be known and $U(p)$ is unknown. Substituting in (88), we have

$$\int_C [Z(p)U(p) - S(p)]e^{pt}\,dp = 0. \tag{91B}$$

This equation will be satisfied if

$$Z(p)U(p) - S(p) = 0, \qquad U(p) = \frac{S(p)}{Z(p)}, \tag{91C}$$

and we again obtain (91).

All that we need, therefore, is the Laplace transform of the independent function. For example, by (80) and (91) the response to the unit step

at $t = \tau$ is

$$A(t-\tau) = \frac{1}{2\pi i}\int_C \frac{e^{p(t-\tau)}}{pZ(p)}\,dp. \tag{92}$$

The response to an upward unit step at $t = \tau - \frac{1}{2}s$ and downward unit step at $t = \tau + \frac{1}{2}s$ is

$$y(t) = \frac{1}{2\pi i}\int_C \frac{2\sinh(ps/2)}{pZ(p)}\,e^{p(t-\tau)}\,dp. \tag{93}$$

If s approaches zero while the magnitude of each step is made equal to $1/s$, we obtain Green's function

$$G(t,\tau) = \frac{1}{2\pi i}\int_C \frac{e^{p(t-\tau)}}{Z(p)}\,dp. \tag{94}$$

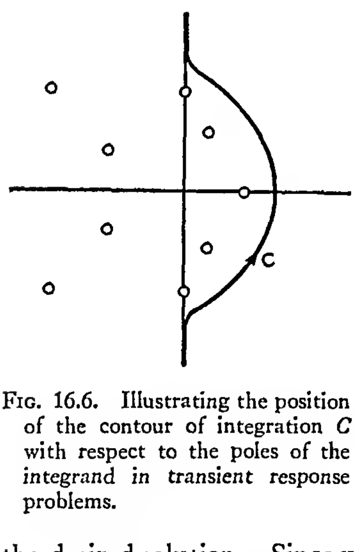

FIG. 16.6. Illustrating the position of the contour of integration C with respect to the poles of the integrand in transient response problems.

In (91) we have a formal solution of (88) in the sense that if we substitute in (88) and differentiate under the integral signs, we shall end up with an identity. However, a question regarding the contour C will arise as soon as we attempt to evaluate (91). In equation (80) for the unit step, C can be deformed in any way as long as the origin is not crossed, and as long as the contour remains along the imaginary axis at infinity; but if this is done in (91) we are likely to cross poles of the integrand other than $p = 0$. In these circumstances (91) may represent several different functions and we are faced with the necessity of choosing C to obtain the desired solution. Since we desire $y(t)$ to equal zero for negative values of t, we should make C pass to the right of all poles of the integrand, Figure 16.6; then closing C by a half-circle at infinity in the right half-plane would leave us with an integral identically equal to zero.

The next question is: are the initial conditions right? In the response to any bounded function $f(t)$, $y(t)$ and its first $(n-1)$ derivatives should vanish at the instant $f(t)$ begins to operate. For the impulse function (which is not bounded but whose integral is) the initial value of the $(n-1)$th derivative should be $1/a_n$. Differentiating (92) $(n-1)$ times,

we have

$$A'(t-\tau) = \frac{1}{2\pi i}\int_C \frac{e^{p(t-\tau)}}{Z(p)}\,dp,$$

$$A''(t-\tau) = \frac{1}{2\pi i}\int_C \frac{pe^{p(t-\tau)}}{Z(p)}\,dp, \tag{95}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$A^{(n-1)}(t-\tau) = \frac{1}{2\pi i}\int_C \frac{p^{n-2}e^{p(t-\tau)}}{Z(p)}\,dp.$$

At $t = \tau$ the typical integral is

$$A^{(m)}(0) = \frac{1}{2\pi i}\int_C \frac{p^{m-1}\,dp}{Z(p)}\,. \tag{96}$$

At infinity $Z(p)$ varies as p^n and the integrand as $1/p^{n+1-m}$; hence, the integral of the absolute value varies as $1/R^{n-m}$. This integral vanishes if $n > m$ and the right-handed closure of C is permissible when $m \leq n - 1$. Since there are no poles inside the augmented contour, the integral vanishes and

$$A^{(m)}(0) = 0, \qquad m \leq n - 1. \tag{97}$$

Green's function is the derivative of the indicial admittance with respect to t; hence

$$G^{(m)}(\tau,\tau) = 0, \qquad m \leq n - 2. \tag{98}$$

The $(n-1)$th derivative of Green's function is

$$G^{(n-1)}(t,\tau) = \frac{1}{2\pi i}\int_C \frac{p^{n-1}e^{p(t-\tau)}}{Z(p)}\,dp. \tag{99}$$

At $t = \tau$, the closure of C is not permissible since the contribution from the infinite semicircle does not vanish. In this case consider the following difference

$$G^{(n-1)}(t,\tau) - a_n^{-1}1(t-\tau) = \frac{1}{2\pi i}\int_C \frac{p^n - a_n^{-1}Z(p)}{pZ(p)}\,e^{p(t-\tau)}\,dp. \tag{100}$$

If $t < \tau$, the integral vanishes and $G^{(n-1)}$ is identical with the second term on the left; thus $G^{(n-1)}(\tau - 0,\tau) = 0$. If $t > \tau$, the left-handed closure of C is permissible. The integral does not normally vanish in this interval. However, we are interested in its value as t approaches $\tau + 0$. The dominant term in the numerator is $-(a_{n-1}/a_n)p^{n-1}$ and, for large p, the integrand varies as $1/p^2$. The contribution from the infinite semicircle is zero, the contour can be returned to its original unclosed form, and then

closed on the right. Hence we find that the integral vanishes at $t = \tau + 0$ and

$$G^{(n-1)}(\tau + 0, \tau) = a_n^{-1} 1(\tau + 0 - \tau) = a_n^{-1}. \tag{101}$$

Thus we have shown that if C is chosen to pass to the right of the zeros of $Z(p)$, the indicial admittance and Green's function for (88) are given by (92) and (94). These formulas reduce the calculation of many responses to a simple routine, particularly when the zeros of $Z(p)$ are simple and the residues can be expressed in terms of the first derivative of $Z(p)$.

The preceding arguments can be generalized to include independent functions other than the unit step and impulse functions. The values of $y(t)$ and its derivatives at the start of $f(t)$ depend on the behavior of $S(p)$ at infinity. If $S(p)$ varies ultimately as $1/p$, $y(t)$ and its first $(n-1)$ derivatives vanish at the start; if $S(p)$ varies as $1/p^2$, the nth derivative also vanishes. The former happens when $f(t)$ is discontinuous at the start; the latter when it is continuous.

5. *Examples of the Laplace transform method*

Let us start with the first example in Section 2 and find the solution of (31). The impedance function is

$$Z(p) = Lp + R. \tag{102}$$

Substituting in (92)

$$A(t) = \frac{1}{2\pi i} \int_C \frac{e^{pt}}{p(Lp + R)}\, dp. \tag{103}$$

The integrand has two poles, $p = 0$ and $p = -R/L$. The residue at $p = 0$ is $1/2\pi iR$ and is obtained as the limit of the integrand multiplied by p as p approaches zero. To obtain the residue at $p = -R/L$ the integrand is multiplied by $p + (R/L)$ and the limit is sought as p approaches $-R/L$; the result is $-(1/2\pi iR)\exp(-Rt/L)$. The integral equals the product of $2\pi i$ and the sum of residues; thus

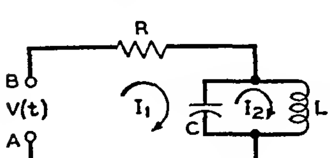

FIG. 16.7. An electric network.

$$A(t) = \frac{1 - \exp(-Rt/L)}{R}. \tag{104}$$

Next let us solve Problem 3 of Section 2. The circuit is shown in Figure 16.7 and the equations are

$$\begin{aligned} RI_1 + \frac{1}{C}\int^t (I_1 - I_2)\, dt &= V(t), \\ \frac{1}{C}\int^t (I_2 - I_1)\, dt + L\frac{dI_2}{dt} &= 0. \end{aligned} \tag{105}$$

Anyone familiar with circuit theory would write directly

$$
\begin{aligned}
Z(p) &= R + \frac{pL(1/pC)}{pL + (1/pC)} = R + \frac{pL}{p^2LC + 1} \\
&= \frac{p^2RLC + pL + R}{p^2LC + 1}.
\end{aligned}
\tag{106}
$$

Suppose, however, that we do not know the simple rules for the construction of impedance functions; then we set

$$
I_1 = \hat{I}_1 e^{pt}, \qquad I_2 = \hat{I}_2 e^{pt}, \qquad V = \hat{V} e^{pt}, \tag{107}
$$

and substitute in (105) to obtain

$$
R\hat{I}_1 + \frac{\hat{I}_1 - \hat{I}_2}{pC} = \hat{V}, \qquad \frac{\hat{I}_2 - \hat{I}_1}{pC} + pL\hat{I}_2 = 0. \tag{108}
$$

The constants of integration have been set equal to zero because we are seeking the steady state response of the form (107).

Solving (108), we have

$$
\hat{I}_1 = \frac{(p^2LC + 1)\hat{V}}{p^2RLC + pL + R}, \qquad \hat{I}_2 = \frac{\hat{V}}{p^2RLC + pL + R}. \tag{109}
$$

Next we let $\hat{V} = S(p)\,dp$ and integrate along the imaginary axis so deformed that there are no poles to the right. To find the indicial admittance we let $\hat{V} = dp/2\pi ip$ so that

$$
A(t) = \frac{1}{2\pi i} \int_C \frac{(p^2LC + 1)e^{pt}\,dp}{p(p^2RLC + pL + R)}. \tag{110}
$$

Let p_1, p_2 be the roots of

$$
p^2RLC + pL + R = 0; \tag{111}
$$

then

$$
p^2RLC + pL + R = RLC(p - p_1)(p - p_2). \tag{112}
$$

To obtain the residues we multiply the integrand successively by $p, p - p_1, p - p_2$ and then let p be $0, p_1, p_2$. Thus we obtain

$$
A(t) = \frac{1}{R} + \frac{e^{p_2t} - e^{p_1t}}{CR^2(p_1 - p_2)}, \qquad t > 0. \tag{113}
$$

It is worth noting that for the roots of (111) we have

$$
p_{1,2}^2 = -(p_{1,2}L + R)/RLC. \tag{114}
$$

The squares and higher powers of the roots of a quadratic equation can always be expressed as linear functions of the roots.

The current $I_2(t)$ in the second mesh can also be calculated from (109).

As a third example we shall consider the response of the circuit consisting of R and L in series when the impressed voltage is

$$
\begin{aligned}
V(t) &= 0, & t &< 0; \\
&= e^{qt}, & 0 &< t < T; \\
&= 0, & t &> T.
\end{aligned}
\tag{115}
$$

The impedance function is given by (102) and we need only the Laplace transform of (115). This we obtain from (86),

$$
S(p) = \frac{1}{2\pi i}\int_0^T e^{(q-p)\tau}\,d\tau = \frac{\exp[(q-p)T]-1}{2\pi i(q-p)}. \tag{116}
$$

Substituting from (102) and (116) in (91), we have

$$
I(t) = \frac{1}{2\pi i}\int_C \frac{\exp[(q-p)T]-1}{(Lp+R)(q-p)}\,e^{pt}\,dp. \tag{117}
$$

There are two poles, at $p = -R/L$ and at $p = q$. In a physical circuit the former is in the left half of the complex plane; but the second pole depends on $V(t)$ and may be anywhere.

At $t = 0$ and $t = T$, $V(t)$ is altered unpredictably and $I(t)$ cannot be represented by a single analytic function. Thus the integral must be evaluated separately in the three regions of definition of $V(t)$ — two, in effect, since C is chosen to make $I(t)$ equal to zero automatically when $t < 0$. In the interval $0 < t < T$, neither left-handed nor right-handed closure is permissible for the entire integral (117). This becomes clear if the integral is separated into two parts,

$$
\begin{aligned}
I(t) = {} & \frac{1}{2\pi i}\int_C \frac{\exp[qT + p(t-T)]}{(Lp+R)(q-p)}\,dp \\
& + \frac{1}{2\pi i}\int_C \frac{\exp(pt)}{(Lp+R)(p-q)}\,dp.
\end{aligned}
\tag{118}
$$

Since $t - T$ is negative, right-handed closure is permissible in the first integral and, since there are no poles inside the contour so formed, the integral vanishes. In the second integral t is positive, left-handed closure is permissible, and by the theorem of residues we find

$$
I(t) = \frac{e^{qt} - e^{-Rt/L}}{Lq + R}, \qquad 0 < t < T. \tag{119}
$$

The resp‹nse is independent of T since prior to $t = T$ the circuit does not

know that the analytic behavior of $V(t)$ is going to be altered at that time.

If $t > T$, then only the left-handed closure is permissible in both integrals. Evaluating the first integral and adding to (119), we have

$$I(t) = \frac{\exp\left[qT - \frac{R}{L}(t - T)\right] - \exp(-Rt/L)}{Lq + R}, \qquad t > T. \quad (120)$$

Thus we have the solution for all q except $q = -R/L$. In this case there is only one pole of the second order and the residues could be calculated as explained in the preceding chapter. It is much simpler, however, to calculate the limits of (119) and (120) as q approaches $-R/L$. Thus for (119) we have

$$\begin{aligned} I(t) &= \frac{e^{-Rt/L}}{Lq + R}\left[\exp\left(q + \frac{R}{L}\right)t - 1\right] \\ &= \frac{e^{-Rt/L}}{Lq + R}\left[1 + \left(q + \frac{R}{L}\right)t + \frac{1}{2}\left(q + \frac{R}{L}\right)^2 t^2 + \cdots - 1\right] \\ &= \frac{1}{L}e^{-Rt/L}t\left[1 + \frac{1}{2}\left(q + \frac{R}{L}\right)t + \cdots\right], \end{aligned} \quad (121)$$

$$\lim I(t) = L^{-1}te^{-Rt/L} \quad \text{as} \quad q \to -R/L.$$

The impedance function may also have zeros of order higher than the first. The best treatment of such cases is similar to the above. Write the solution for the case of simple poles and let the required number of these poles approach each other. The limits of the indeterminate forms are then obtained by the usual methods, such as successive differentiation of the numerator and denominator.

Problems

1. Obtain the answers to the problems in Sections 2 and 3 by the Laplace transform method.

2. Find the Laplace transform of

$$\begin{aligned} f(t) &= 0, & t &< 0; \\ &= t/T, & 0 &< t < T; \\ &= 1, & t &> T. \end{aligned}$$

Ans. $$S(p) = \frac{1 - e^{-pT}}{2\pi i T p^2}.$$

3. Find the Laplace transform of the following trapezoidal function

$$\begin{aligned} f(t) &= 0, \qquad t < 0; \\ &= t/T, \qquad 0 < t < T; \\ &= 1, \qquad T < t < T + T_1; \\ &= (2T + T_1 - t)/T, \qquad T + T_1 < t < 2T + T_1; \\ &= 0, \qquad t > 2T + T_1. \end{aligned}$$

Ans. $$S(p) = \frac{1 - e^{-pT} - e^{-p(T+T_1)} + e^{-p(2T+T_1)}}{2\pi i T p^2}.$$

4. Find the Laplace transform of

$$\begin{aligned} f(t) &= 0, \qquad t < 0; \\ &= \cos(\omega t + \varphi), \qquad 0 < t < T; \\ &= 0, \qquad t > T. \end{aligned}$$

Ans. $$S(p) = \frac{1}{4\pi i}\left[\frac{1 - e^{-(p-i\omega)T}}{p - i\omega} e^{i\varphi} + \frac{1 - e^{-(p+i\omega)T}}{p + i\omega} e^{-i\varphi}\right].$$

Note that $$S(p) = S_1(p,\varphi) - e^{-pT} S_1(p, \varphi + \omega T),$$

where $$S_1(p,\varphi) = \frac{1}{4\pi i}\left(\frac{e^{i\varphi}}{p - i\omega} + \frac{e^{-i\varphi}}{p + i\omega}\right)$$

$$= \frac{p\cos\varphi - \omega \sin\varphi}{2\pi i(p^2 + \omega^2)}.$$

The function $S_1(p,\varphi)$ is independent of T and is the Laplace transform of the sinusoid starting at $t = 0$ and continuing indefinitely.

5. Solve $L(dI/dt) + RI = f(t)$ where $f(t)$ is defined as in problems 2, 3, 4. Verify the answer by direct substitution in the differential equation; show directly that I and its first derivative vanish at $t = 0$.

6. Replace the differential equation in Problem 5 by

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int^t I\, dt = f(t)$$

and solve the new set of problems.

7. Obtain equations (119) and (120) directly from the differential equation by writing down the solutions in the three intervals defined in (115) as sums of the steady state and transient solutions and then matching them at points of discontinuity.

8. Show that $I(t)$ given by (117) vanishes when $t < 0$ even if C passes to the left of $p = q$, the pole introduced by $V(t)$. Show that this alteration in the contour of integration does not affect either (119) or (120).

6. *The Heaviside expansion theorems*

If the zeros of $Z(p)$ are simple, then in the vicinity of each zero $Z(p) = (p - p_m)Z'(p_m)$ and by the theorem of residues (94) becomes (if $\tau = 0$)

$$G(t,0) = \sum_m \frac{e^{p_m t}}{Z'(p_m)}, \qquad t > 0, \tag{122}$$

the summation being extended over all zeros of $Z(p)$.

Similarly for the indicial admittance we have

$$A(t) = \frac{1}{Z(0)} + \sum_m \frac{e^{p_m t}}{p_m Z'(p_m)}, \qquad t > 0, \tag{123}$$

provided $Z(0) \neq 0$. If $Z(0) = 0$, then the first term becomes $t/Z'(0)$; provided, of course, $Z'(0) \neq 0$.

7. *Arbitrary initial conditions*

So far we have been concerned with that particular solution of a set of differential equations for a given system which can be interpreted as the response to a given set of forces impressed on the system. These forces were supposed to be impressed at some instant $t = 0$; and, in order to conform to the physical behavior of the system, the solution had to vanish when $t < 0$.

In other types of problems the state of the system at some instant $t = 0$ is given, and it is required to find its state at any subsequent instant, assuming that there are no impressed forces when $t > 0$. The solution will represent the response of the system to those impressed forces which have ceased at $t = 0$. There are two basic approaches to the solution of such problems. Since there are no impressed forces when $t > 0$, our system of differential equations is homogeneous. We may look for its general solution and then try to determine the arbitrary constants so as to satisfy the given initial conditions. This method was explained in Chapter 12. The second method is based on the possibility of replacing all impressed forces which have been acting up to $t = 0$ by a set of equivalent forces impressed at $t = 0$. When this is done, the problem becomes identical with that considered in the preceding sections of this chapter.

Suppose, for instance, that we want that solution of

$$a_2 \frac{d^2y}{dt^2} + a_1 \frac{dy}{dt} + a_0 y = 0, \tag{124}$$

for which y and its derivative are equal to $y(0)$ and $y'(0)$ at $t = 0$. In

Chapter 11 we found such a solution

$$\begin{aligned} y(t) &= y(0)\frac{p_2e^{p_1t} - p_1e^{p_2t}}{p_2 - p_1} + y'(0)\frac{e^{p_2t} - e^{p_1t}}{p_2 - p_1}, \\ a_2p_{1,2}^2 &+ a_1p_{1,2} + a_0 = 0. \end{aligned} \tag{125}$$

On the other hand, if we make $y(t) = 0$ for $t < 0$, the initial values $y(0)$, $y'(0)$ become upward discontinuities in the function and its derivative. Hence we should be able to obtain $y(t)$ as the sum of responses to certain impulse functions. For this purpose we rewrite (124) as a system of first-order equations by introducing another dependent function,

$$y_1(t) = \frac{dy(t)}{dt}. \tag{126}$$

The system is

$$\frac{dy}{dt} = y_1, \qquad a_2\frac{dy_1}{dt} = -a_1y_1 - a_0y. \tag{127}$$

To obtain that solution which equals zero when $t < 0$ and has the required upward discontinuities, we must solve

$$\begin{aligned} \frac{dy}{dt} &= y_1 + y(0)I(t,0), \\ a_2\frac{dy_1}{dt} &= -a_1y_1 - a_0y + a_2y'(0)I(t,0). \end{aligned} \tag{128}$$

To prove this we integrate (128) over $(-0,+0)$.

$$\begin{aligned} y(+0) - y(-0) &= \int_{-0}^{+0} y_1\,dt + y(0)\int_{-0}^{+0} I(t,0)\,dt = y(0), \\ a_2[y_1(+0) - y_1(-0)] &= a_2y'(0). \end{aligned} \tag{129}$$

To solve (128) by the Laplace transform method, we write

$$\begin{aligned} y(t) &= \int_C S(p)e^{pt}\,dp, \qquad y_1(t) = \int_C S_1(p)e^{pt}\,dp, \\ I(t,0) &= \int_C T(p)e^{pt}\,dp. \end{aligned} \tag{130}$$

Presently we shall consider the nature of $T(p)$; but first let us follow through the formal procedure. Substituting from (130) in (128), performing the necessary differentiations, and transferring all terms to the

left side, we have

$$\int_C [pS(p) - S_1(p) - y(0)T(p)]e^{pt}\,dp = 0,$$
$$\int_C [(a_2p + a_1)S_1(p) + a_0S(p) - y'(0)T(p)]e^{pt}\,dp = 0. \tag{131}$$

These equations can be satisfied if we equate the bracketed expressions to zero; thus

$$\begin{aligned} pS(p) - S_1(p) &= y(0)T(p), \\ a_0S(p) + (a_2p + a_1)S_1(p) &= a_2y'(0)T(p). \end{aligned} \tag{132}$$

The new equations may now be solved for S and S_1; the solution for S will be sufficient since $y'(t)$ may subsequently be obtained by differentiating $y(t)$. Thus

$$S(p) = \frac{(a_2p + a_1)y(0) + a_2y'(0)}{a_2p^2 + a_1p + a_0}\,T(p), \tag{133}$$

$$y(t) = \int_C \frac{(a_2p + a_1)y(0) + a_2y'(0)}{a_2p^2 + a_1p + a_0}\,e^{pt}T(p)\,dp. \tag{134}$$

Now let us return to $T(p)$. The impulse function is the limit of two step functions, of magnitudes $1/s$ and $-1/s$, the first starting at $t = -s/2$ and the second at $t = s/2$. For the unit impulse of finite duration s, we have therefore

$$I(t;-\tfrac{1}{2}s, \tfrac{1}{2}s) = \frac{1}{\pi i}\int_C \frac{\sinh(\frac{1}{2}ps)}{sp}\,e^{pt}\,dp. \tag{135}$$

Thus the Laplace transform of an impulse of *finite* duration is

$$T(p) = \frac{\sinh(\frac{1}{2}ps)}{\pi isp}. \tag{136}$$

As s approaches zero we have in the limit

$$T(p) = 1/2\pi i. \tag{137}$$

However, if we pass to the limit in (135), we shall be left with a divergent integral. The Laplace transform of an infinitely short impulse does not exist and it might seem that our method becomes inoperative; but it does not. We do not have to use infinitely short impulse functions in (128); impulses of finite duration would serve our purpose just as well. Instead of forcing sudden discontinuities in $y(t)$ and $y'(t)$ at $t = 0$, we can make

$y(t)$ and $y'(t)$ change rapidly, as rapidly as we please, when passing $t = 0$. Thus we can use (136) in (134) and then let s approach zero. The limit,

$$y(t) = \frac{1}{2\pi i}\int_C \frac{(a_2p + a_1)y(0) + a_2y'(0)}{a_2p^2 + a_1p + a_0} e^{pt}\, dp, \tag{138}$$

exists and yields the desired solution of (128).

It should be noted that if $S(p)$ and $S_1(p)$ satisfy (132), equations (131) are satisfied regardless of C; the choice of C is dictated by the condition that $y(t)$ and $y'(t)$ must vanish for $t < 0$. If we make C pass to the left of the poles of $S(p)$, we shall have a solution whose *end values* are $-y(0)$, $-y'(0)$; the discontinuities in $y(t)$ and $y'(t)$ are the same as for the former contour. As seen from (129), equations (128) imply specified discontinuities; it is left to the contour C to discriminate between several possible ways in which these discontinuities may take place.

If we evaluate (138) by the theorem of residues, we find an answer somewhat different from (125); however, the two forms may be reconciled if we use the identity $p_1 + p_2 = -a_1/a_2$.

So far we have seemingly gone to a great deal of trouble to solve a simple problem. High power methods are seldom seen to best advantage in elementary examples. The Laplace transform method shows itself to better advantage if applied to higher order equations. The direct method of writing the general solution, then solving a system of linear algebraic equations to find the arbitrary constants, is simple in principle and the solutions are readily expressed in terms of determinants; but the evaluation of these determinants is another matter. The Laplace transform method yields a simpler general solution. The method becomes particularly valuable in the case of partial differential equations.

The following is the procedure for solving the nth order equation (88), made homogeneous by letting $f(t)$ vanish. Introduce new variables and rewrite the equation as follows:

$$\begin{gathered} y = y_0, \quad \frac{dy_0}{dt} = y_1, \quad \frac{dy_1}{dt} = y_2, \quad \cdots, \quad \frac{dy_{n-2}}{dt} = y_{n-1}, \\ a_n \frac{dy_{n-1}}{dt} + a_{n-1}y_{n-1} + a_{n-2}y_{n-2} + \cdots + a_1y_1 + a_0y_0 = 0. \end{gathered} \tag{139}$$

Replace this set of equations by

$$\begin{gathered} \frac{dy_m}{dt} = y_{m+1} + y^{(m)}(0)I(t,0), \qquad m = 0, 1, \cdots n-2; \\ a_n \frac{dy_{n-1}}{dt} + \sum_{m=0}^{n-1} a_m y_m = a_n y^{(n-1)}(0)I(t,0), \end{gathered} \tag{140}$$

where $y_0 = y$ and $y^{(m)}(0)$ is the initial value of the mth derivative. Express each $y_m(t)$ as a Laplace integral,

$$y_m(t) = \int_C S_m(p)e^{pt}\,dp, \tag{141}$$

and substitute in (140). Transfer all terms on one side of each equation and equate the integrands to zero

$$\begin{aligned} S_{m+1} &= pS_m - y^{(m)}(0)T, \qquad m \leq n-2; \\ a_npS_{n-1} + \sum_{m=0}^{n-1} a_mS_m &= a_ny^{(n-1)}(0)T. \end{aligned} \tag{142}$$

These equations may be solved successively; thus

$$\begin{aligned} S_1 &= pS_0 - y(0)T, \\ S_2 &= p^2S_0 - py(0)T - y'(0)T, \\ S_3 &= p^3S_0 - p^2y(0)T - py'(0)T - y''(0)T, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_{n-1} &= p^{n-1}S_0 - p^{n-2}y(0)T - p^{n-3}y'(0)T - \cdots - y^{(n-2)}(0)T. \end{aligned} \tag{143}$$

Substitute these results in the last equation of the set (142) and collect the terms containing S_0, $y(0)$, $y'(0)$, $\cdots$. A certain regularity will be observed in the formation of the coefficients; thus it will be found that

$$Z(p)S_0 = [Z_1(p)y(0) + Z_2(p)y'(0) + \cdots + Z_n(p)y^{(n-1)}(0)]T, \tag{144}$$

where

$$\begin{aligned} Z(p) &= a_np^n + a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \cdots + a_2p^2 + a_1p + a_0, \\ Z_1(p) &= a_np^{n-1} + a_{n-1}p^{n-2} + a_{n-2}p^{n-3} + \cdots + a_2p + a_1, \\ Z_2(p) &= a_np^{n-2} + a_{n-1}p^{n-3} + a_{n-2}p^{n-4} + \cdots + a_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Z_n(p) &= a_n \end{aligned} \tag{145}$$

The first of these functions, $Z(p)$, is the impedance function, already encountered in (90); each of the following functions is obtained from the preceding by dropping the last term and dividing the result by p. The Laplace transform of $y_0(t) = y(t)$,

$$S_0 = \frac{Z_1(p)y(0) + Z_2(p)y'(0) + \cdots + Z_n(p)y^{(n-1)}(0)}{2\pi iZ(p)}, \tag{146}$$

is obtained from (144); thus we have the required solution

$$y_0(t) = y(t) = \int_C S_0(p)e^{pt}\,dp. \tag{147}$$

If the zeros of $Z(p)$ are simple, then

$$\begin{aligned} y(t) = y(0) \sum \frac{Z_1(p_m)}{Z'(p_m)} e^{p_m t} + y'(0) \sum \frac{Z_2(p_m)}{Z'(p_m)} e^{p_m t} + \cdots \\ + y^{(n-1)}(0) \sum \frac{Z_n(p_m)}{Z'(p_m)} e^{p_m t}, \end{aligned} \tag{148}$$

where the summation is extended over all zeros.

Problems

1. Apply (148) to (124) and compare with (125). Note that $a_1/a_2 = -(p_1 + p_2)$

2. Solve $\dfrac{d^2y}{dt^2} - z = 0$, $\dfrac{dz}{dt} + 2z - y = 0$, $y(0) = 1$, $y'(0) = 0$, $z(0) = 1$.

Ans.

$$y(t) = 0.1(5 + \sqrt{5})e^{p_1 t} + 0.1(5 - \sqrt{5})e^{p_2 t},$$

$$z(t) = 0.1(5 - \sqrt{5})e^{p_1 t} + 0.1(5 + \sqrt{5})e^{p_2 t},$$

$$p_{1,2} = \tfrac{1}{2}(-1 \pm \sqrt{5}).$$

3. Solve $\dfrac{d^2x}{dt^2} + x + y = 0$, $\dfrac{d^2y}{dt^2} - x + y = 0$, subject to arbitrary initial conditions.

Ans. The Laplace transforms are

$$\bar{x} = \frac{(1 + p^2)[x'(0) + px(0)] - [y'(0) + py(0)]}{2\pi i[(1 + p^2)^2 + 1]},$$

$$\bar{y} = \frac{(1 + p^2)[y'(0) + py(0)] + [x'(0) + px(0)]}{2\pi i[(1 + p^2)^2 + 1]}.$$

If $\qquad a = \sqrt[4]{2}\exp(3\pi i/8),$

$$\begin{aligned} x(t) = \tfrac{1}{2}x(0)(\cosh at + \cosh a^*t) + \tfrac{1}{2}x'(0)\left(\frac{\sinh at}{a} + \frac{\sinh a^*t}{a^*}\right) \\ + \tfrac{1}{2}iy(0)(\cosh at - \cosh a^*t) + \tfrac{1}{2}iy'(0)\left(\frac{\sinh at}{a} - \frac{\sinh a^*t}{a^*}\right), \end{aligned}$$

$$\begin{aligned} y(t) = \tfrac{1}{2}y(0)(\cosh at + \cosh a^*t) - \tfrac{1}{2}ix(0)(\cosh at - \cosh a^*t) \\ + \tfrac{1}{2}y'(0)\left(\frac{\sinh at}{a} + \frac{\sinh a^*t}{a^*}\right) - \tfrac{1}{2}ix'(0)\left(\frac{\sinh at}{a} - \frac{\sinh a^*t}{a^*}\right). \end{aligned}$$

8. *A simple problem of wave propagation*

When the independent variable is a spatial coordinate, the boundary conditions may resemble those treated in the preceding section, but different sets of conditions may also occur. First, let us consider waves in an infinitely long nondissipative electrical transmission line, energized by a series generator of zero impedance at $x = 0$, Figure 16.8. The equations to be solved are

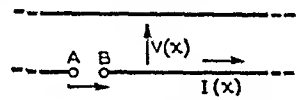

FIG. 16.8. An electric transmission line.

$$\frac{dV}{dx} = -i\omega LI + V_0 I(x,0), \qquad \frac{dI}{dx} = -i\omega CV, \tag{149}$$

where V_0 is the electromotive force developed by the generator and ω is the frequency in radians per second. The impressed force appears as a discontinuity, $V(+0) - V(-0) = V_0$, in the voltage across the line; the longitudinal current is continuous.

Expressing $V(x)$, $I(x)$ as Laplace integrals,

$$V(x) = \int_C S_1(\gamma)e^{\gamma x}\, d\gamma, \qquad I(x) = \int_C S_2(\gamma)e^{\gamma x}\, d\gamma, \tag{150}$$

substituting in (149), and rearranging the terms,

$$\begin{aligned} \int_C [\gamma S_1(\gamma) + i\omega LS_2(\gamma) - V_0 T(\gamma)]e^{\gamma x}\, d\gamma = 0, \\ \int_C [\gamma S_2(\gamma) + i\omega CS_1(\gamma)]e^{\gamma x}\, d\gamma = 0. \end{aligned} \tag{151}$$

These equations are satisfied if

$$\gamma S_1 + i\omega LS_2 = V_0 T, \qquad i\omega CS_1 + \gamma S_2 = 0. \tag{152}$$

Solving,

$$S_1(\gamma) = \frac{\gamma V_0 T(\gamma)}{\gamma^2 + \omega^2 LC}, \qquad S_2(\gamma) = -\frac{i\omega CV_0 T(\gamma)}{\gamma^2 + \omega^2 LC}; \tag{153}$$

substituting in (150) and letting $T = 1/2\pi i$,

$$V(x) = \frac{V_0}{2\pi i}\int_C \frac{\gamma e^{\gamma x}\, d\gamma}{\gamma^2 + \omega^2 LC}, \qquad I(x) = -\frac{\omega CV_0}{2\pi}\int_C \frac{e^{\gamma x}\, d\gamma}{\gamma^2 + \omega^2 LC}. \tag{154}$$

Thus we have obtained a solution in which $I(x)$ is continuous and $V(x)$ increases by V_0 in passing through $x = 0$ from left to right. However, the solution is still not unique, since the contour C has not been specified. It is here that physical considerations again enter the scene. The required

solution must be such that the phase of the response is delayed with increasing distance from the source; this condition requires C to go round the poles of the integrands as shown in Figure 16.9. If $x > 0$, the left-handed closure of C is permissible and the response is proportional to $\exp(-i\omega x\sqrt{LC})$; if $x < 0$, the right-handed closure is permissible and the response is proportional to $\exp(i\omega x\sqrt{LC})$. Thus, the behavior of the response at $x = \pm\infty$ is correct.

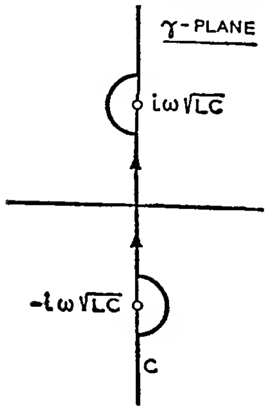

FIG. 16.9. Illustrating the position of the contour of integration with respect to the poles of the integrand in spatial response problems.

If the transmission line is dissipative, there will be a resistive term RI in the first equation in (149) and a conductive term in the second equation. The poles are then at $\gamma = \pm\sqrt{(R + i\omega L)(G + i\omega C)}$. In transmission lines R, L, G, C are positive; hence, one of these poles is in the first quadrant and the other in the third. If C is taken along the imaginary axis, it will separate the poles in such a way that the answer will be correct; the response will decrease with the distance from the source. As the dissipation becomes smaller, the poles move nearer to the imaginary axis and in the limit the contour C is indented.

The point generator may be replaced by any series voltage distribution. If $E(x)$ is the impressed voltage per unit length and if $S(\gamma)$ is its Laplace transform,

$$V(x) = \int_C \frac{\gamma S(\gamma)e^{\gamma x}\,d\gamma}{\gamma^2 + \omega^2 LC}, \qquad I(x) = -\,i\omega C\int_C \frac{S(\gamma)e^{\gamma x}\,d\gamma}{\gamma^2 + \omega^2 LC}, \tag{155}$$

where C should separate the poles of the integrand which belong to the system, and not to $S(\gamma)$, as shown in Figure 16.9. The poles of $S(\gamma)$ may be placed on either side of C.

Suppose now that the transmission line is finite, beginning at $x = 0$ and ending at $x = \ell$. Can we ask for the solution in which the initial values are given: $V(0) = V_0$, $I(0) = I_0$? Certainly. We can add $I_0 I(x,0)$ in the second equation in (149) and choose C exactly as in the preceding section, since the problem is the same. The considerations involving the choice of C are now purely mathematical; physical conditions, aside from the transmission line itself, have not yet entered into the problem; they will have to be imposed when we specify V_0, I_0. If the line is electrically open at $x = 0$, then $I_0 = 0$; the solution will then tell us the voltage $V(\ell)$ which must be maintained at $x = \ell$ in order that the voltage

across the open end may be V_0. Alternatively, if we know $V(\ell)$, the solution will tell us V_0. If there is a resistance R across the line at $x = 0$, then, with the convention concerning the positive directions of V and I implied in (149), we should have $V_0 = -RI_0$. If $V_0 = RI_0$, we automatically assume that at $x = 0$ there is a generator, and that this generator "sees" the line as a resistance R. The ratio $V(\ell)/I(\ell)$ in the corresponding solution will be independent of I_0 and will give the impedance needed across the line at $x = \ell$ in order to satisfy the assumption at $x = 0$.

9. *Waves between parallel planes*

The one-dimensional problem of the preceding section is too simple to show the advantages of the Laplace transform method. Let us see how the method works in the case of waves between two parallel, perfectly conducting planes, whose cross section is shown in Figure 16.10. We assume an a-c voltage acting between two half-planes from $x = -s/2$ to $x = s/2$. Let this voltage distribution be independent of z. This voltage might be delivered by a wave traveling along another pair of parallel planes so close together that the electric lines are forced to be straight and perpendicular to the guiding planes, except in the immediate vicinity of the junction. We shall assume that over the "gap" E_x is known; elsewhere at the plane $y = 0$, E_x vanishes since we have assumed perfect conductivity. Thus we have

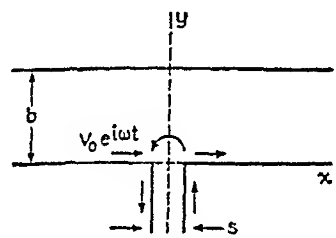

Fig. 16.10. A cross section of two pairs of parallel planes.

$$
\begin{aligned}
E_x e^{i\omega t} &= E_0(x) e^{i\omega t}, \qquad -s/2 < x < s/2, \qquad y = 0; \\
&= 0, \qquad\qquad\qquad |x| > s/2, \qquad\quad y = 0; \qquad (156)
\end{aligned}
$$

$$
\int_{-s/2}^{s/2} E_0(x)\, dx = V_0.
$$

The voltage V_0 is the field voltage across the gap; it is equal and opposite to the impressed voltage. If $S_0(\gamma)$ is the Laplace transform of $E_0(x)$, then *at the plane* $y = 0$,

$$
E_x = \int_C S_0(\gamma) e^{\gamma x}\, d\gamma, \qquad S_0(\gamma) = \frac{1}{2\pi i} \int_{-s/2}^{s/2} E_0(x) e^{-\gamma x}\, dx. \qquad (157)
$$

The differential equations governing the field elsewhere are*

$$\frac{\partial H_z}{\partial y} = i\omega\epsilon E_x, \qquad \frac{\partial H_z}{\partial x} = -i\omega\epsilon E_y, \qquad \frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} = i\omega\mu H_z. \tag{158}$$

The problem is to find that solution of these equations which matches the given field at $y = 0$. The matching is facilitated if H_z, E_x, E_y are expressed as Laplace integrals:

$$H_z = \int_C h_z(\gamma,y)e^{\gamma x}\,d\gamma, \qquad E_x = \int_C e_x(\gamma,y)e^{\gamma x}\,d\gamma,$$

$$E_y = \int_C e_y(\gamma,y)e^{\gamma x}\,d\gamma; \tag{159}$$

thus

$$e_x(\gamma,0) = S_0(\gamma), \qquad e_x(\gamma,b) = 0. \tag{160}$$

The second equation comes from the vanishing of the tangential electric vector at the perfectly conducting plane $y = b$.

Substituting (159) in (158),

$$\int_C \left(\frac{\partial h_z}{\partial y} - i\omega\epsilon e_x\right) e^{\gamma x}\,d\gamma = 0, \qquad \int_C (\gamma h_z + i\omega\epsilon e_y)e^{\gamma x}\,d\gamma = 0,$$

$$\int_C \left(\frac{\partial e_x}{\partial y} - \gamma e_y - i\omega\mu h_z\right) e^{\gamma x}\,d\gamma = 0. \tag{161}$$

These equations are satisfied if the integrands are equal to zero

$$\frac{\partial h_z}{\partial y} = i\omega\epsilon e_x, \qquad i\omega\epsilon e_y = -\gamma h_z, \qquad \frac{\partial e_x}{\partial y} - \gamma e_y = i\omega\mu h_z. \tag{162}$$

The symbols of partial differentiation have been retained because h_z, e_x, e_y are functions of γ as well as of y; but γ is only a parameter, and not an independent variable, in (162). Consequently, the Laplace transform method has enabled us to reduce a system of partial differential equations containing two independent variables to a system of ordinary differential equations. To solve the latter, eliminate e_x, e_y:

$$\frac{\partial^2 h_z}{\partial y^2} = (-\gamma^2 - \omega^2\mu\epsilon)h_z; \tag{163}$$

* S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand Co., New York, 1943, p. 95, equation (12–5).

hence

$$h_z = A \cosh (y - b) \sqrt{-\gamma^2 - \omega^2\mu\epsilon} + B \sinh (y - b) \sqrt{-\gamma^2 - \omega^2\mu\epsilon}. \tag{164}$$

This form of solution is adopted in anticipation of the effect of the boundary condition (160) at $y = b$. From (162) and (164), we have

$$e_x = \frac{\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}{i\omega\epsilon} [A \sinh (y - b) \sqrt{-\gamma^2 - \omega^2\mu\epsilon} + B \cosh (y - b) \sqrt{-\gamma^2 - \omega^2\mu\epsilon}]. \tag{165}$$

Since e_x vanishes at $y = b$, we must have $B = 0$; and to satisfy the condition at $y = 0$, we must have

$$-A \frac{\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}{i\omega\epsilon} \sinh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon} = S_0(\gamma),$$
$$A = - \frac{i\omega\epsilon S_0(\gamma)}{\sqrt{-\gamma^2 - \omega^2\mu\epsilon} \sinh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon}}. \tag{166}$$

Thus we have obtained the Laplace transforms of the electric and magnetic vectors and the rest of the problem will consist of the evaluation of (159).

Let us assume that $s = 0$, and evaluate H_z at $y = 0$. For $s = 0$, $S_0(\gamma) = V_0/2\pi i$ and

$$H_z = - \frac{\omega\epsilon V_0}{2\pi} \int_C \frac{\cosh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon}}{\sqrt{-\gamma^2 - \omega^2\mu\epsilon} \sinh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon}} e^{\gamma x}\, d\gamma. \tag{167}$$

Let x be positive so that the left-handed closure of the contour is permissible. The poles of the integrand are the zeros of the denominator,

$$\sqrt{-\gamma^2 - \omega^2\mu\epsilon} \sinh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon} = 0. \tag{168}$$

Since $\sinh u$ vanishes when $u = in\pi$, we have

$$b \sqrt{-\gamma^2 - \omega^2\mu\epsilon} = in\pi, \quad n = 0, \pm 1, \pm 2, \cdots,$$
$$\gamma = \pm \sqrt{\frac{n^2\pi^2}{b^2} - \omega^2\mu\epsilon} = \pm \Gamma_n, \tag{169}$$

where *by definition* Γ_n is either positive real or positive imaginary, depending on whether $n\pi$ is greater or less than $\omega b\sqrt{\mu\epsilon}$. Since

$$\frac{d}{d\gamma} \sinh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon} = - \frac{\gamma b}{\sqrt{-\gamma^2 - \omega^2\mu\epsilon}} \cosh b \sqrt{-\gamma^2 - \omega^2\mu\epsilon} \tag{170}$$

does not vanish at $\gamma = \pm\Gamma_n$, $n > 0$, the poles are simple. In the exceptional case, $n = 0$, $\sinh b\sqrt{-\gamma^2 - \omega^2\mu\epsilon} = b\sqrt{-\gamma^2 - \omega^2\mu\epsilon}$ and $\gamma = \pm i\omega\sqrt{\mu\epsilon}$; these poles are also simple. By the theorem of residues

$$H_z = -\frac{1}{2b}\sqrt{\epsilon/\mu}V_0\, e^{-i\omega\sqrt{\mu\epsilon}x} - \frac{i\omega\epsilon V_0}{b}\sum_{n=1}^{\infty}\frac{1}{\Gamma_n}e^{-\Gamma_n x}. \tag{171}$$

This series converges for all positive values of x. As $x \to 0$, the distant terms of the series vary as $1/n$; thus the series will diverge at $x = 0$. If the length s of the gap is different from zero, then the corresponding series converges for *all* values of x.

For any finite value of b, a finite number of Γ_n's are imaginary. These terms represent traveling waves. The remaining waves are strictly local, being exponentially attenuated with increasing distance from the source. If $\omega\sqrt{\mu\epsilon}b < \pi$, only the first term in (171) represents a traveling wave.

10. *Branch points*

The convergence of the series (171) deteriorates as b increases; the individual terms become smaller, the Γ_n's are crowded together and a large number of terms must be used. Instead of trying to find the limit of the series as $b \to \infty$, it is easier to solve the problem by starting with $b = \infty$. The procedure is identical with that given in the preceding section; but in the present case (164) should be written as follows:

$$h_z = A \exp y\sqrt{-\gamma^2 - \omega^2\mu\epsilon}. \tag{172}$$

The reason for the single term lies in the absence of reflected waves. At $y = \infty$, the phase of the wave should be progressively retarded and the intensity should be finite. These physical considerations settle for us the sign of the square root in (172) and the position of C in relation to the branch points $\gamma = \pm i\omega\sqrt{\mu\epsilon}$. Since γ is imaginary on C (except at the indentations round the branch points), we let $\gamma = i\chi$; hence $\sqrt{-\gamma^2 - \omega^2\mu\epsilon} = \sqrt{\chi^2 - \omega^2\mu\epsilon}$. Between the branch points the square root is imaginary; it must be negative if the phase of (172) is to be retarded with increasing y. Hence,

$$\sqrt{-\gamma^2 - \omega^2\mu\epsilon} = -i\sqrt{\omega^2\mu\epsilon - \chi^2}, \qquad |\chi| < \omega\sqrt{\mu\epsilon}. \tag{173}$$

Outside this interval the square root is real; it must be negative or else the field would become infinite at $y = \infty$. Thus

$$\sqrt{-\gamma^2 - \omega^2\mu\epsilon} = -\sqrt{\chi^2 - \omega^2\mu\epsilon}, \qquad |\chi| > \omega\sqrt{\mu\epsilon}. \tag{174}$$

In order to connect these values continuously on C, the indentations should

be to the right of $\gamma = -i\omega\sqrt{\mu\epsilon}$ and to the left of $\gamma = i\omega\sqrt{\mu\epsilon}$. From now on we proceed as in the previous section and obtain

$$\begin{gathered} e_z = \frac{A\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}{i\omega\epsilon} \exp y\sqrt{-\gamma^2 - \omega^2\mu\epsilon}, \\ \frac{A\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}{i\omega\epsilon} = S_0(\gamma), \qquad A = \frac{i\omega\epsilon S_0(\gamma)}{\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}. \end{gathered} \tag{175}$$

For a finite voltage across an infinitesimal gap, we have

$$H_z = \frac{i\omega\epsilon V_0}{2\pi i}\int_C \frac{\exp(y\sqrt{-\gamma^2 - \omega^2\mu\epsilon} + \gamma x)}{\sqrt{-\gamma^2 - \omega^2\mu\epsilon}}\, d\gamma. \tag{176}$$

The calculation of this integral is not easy. The integrand is multiple-valued and it has no poles. The contributions to the integral from the infinitely small indentations are nil; but indirectly they affect the value of the integral by determining the relative signs of the integrand on different parts of C. If the integrand were single-valued, the integral would equal zero (since there are no poles); as it is, we have

$$\begin{gathered} H_z = -i\omega\epsilon\pi^{-1}V_0K_0(i\beta\rho) = -\tfrac{1}{2}\omega\epsilon V_0[J_0(\beta\rho) - iN_0(\beta\rho)], \\ \rho = \sqrt{x^2 + y^2}, \qquad \beta = \omega\sqrt{\mu\epsilon}. \end{gathered} \tag{177}$$

This value can be obtained more easily by solving the problem in cylindrical coordinates than by evaluating (176).

The moral of this example is: *beware of branch points.* Do not treat the integrals of multiple-valued functions as you would the integrals of single-valued functions; give them special attention.

11. *Waves on an infinite cylinder*

In the case of propagation of electric charge on the surface of an infinitely long perfectly conducting cylinder there are several possibilities: We may be interested in propagation on the external surface as illustrated by Figures 16.11(a,d), or on the internal surface as shown in Figures 16.11(b,e), or on both surfaces as in Figure 16.11c. In Figures 16.11(a,b,c) it is assumed that we know the longitudinal field across the gap between the cylinders; in Figures 16.11(d,e) the exciting field is generated by an ideal solenoid or a "magnetic current loop."

In all these cases the current in the cylinder is

$$I(z) = \int_C Y(\gamma)T(\gamma)e^{\gamma z}\, d\gamma, \tag{178}$$

where $T(\gamma)$ depends solely on the exciting field and $Y(\gamma)$ is determined by the cylinder and the regions exposed to the excitation. Thus in the cases shown in Figures 16.11 (a,b,c) we have successively

$$Y(\gamma) = Y_e(\gamma) = \frac{i\beta a K_1(a\sqrt{-\gamma^2-\beta^2})}{60\sqrt{-\gamma^2-\beta^2}K_0(a\sqrt{-\gamma^2-\beta^2})},$$

$$Y(\gamma) = Y_i(\gamma) = \frac{i\beta a I_1(a\sqrt{-\gamma^2-\beta^2})}{60\sqrt{-\gamma^2-\beta^2}I_0(a\sqrt{-\gamma^2-\beta^2})}, \tag{179}$$

$$Y(\gamma) = Y_e(\gamma) + Y_i(\gamma),$$

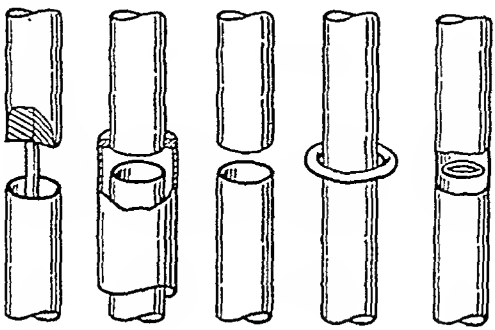

FIG. 16.11. Illustrating different ways of electric excitation of a metal cylinder: (a) the power is delivered by a coaxial transmission line to an opening in the outer cylinder to excite external waves on that cylinder; (b) a similar excitation of internal waves inside the inner cylinder; (c) excitation of internal and external waves assuming that electric charge is carried back and forth between the two halves of the cylinder; (d) excitation of external waves by means of a solenoid; (e) similar excitation of internal waves.

where a is the radius of the cylinder, $\beta = 2\pi/\lambda$, and λ is the wavelength. The cases in Figures 11.16(a,d) differ only in the form of $T(\gamma)$; similarly the cases in Figures 11.16(b,e) have the same $Y(\gamma)$.

Internal and external waves differ profoundly. The admittance function $Y_i(\gamma)$ has no branch points. At first it may look as if $\gamma = \pm i\beta$ were branch points; but $I_1(u)/u$ and $I_0(u)$ are even functions of u and the square roots disappear from them. The integral (178) may thus be expressed as a series of terms corresponding to the zeros of $I_0(a\sqrt{-\gamma^2-\beta^2})$. Each term will represent what is known as a "mode of transmission." The propagation constants, γ, of these modes are determined by the above zeros. In the case of external waves the integrand has branch points but no poles; there is no discrete set of transmission modes.

For an infinitely narrow gap we have, of course, $T(\gamma) = V^i/2\pi i$, where V^i

is the impressed voltage; the corresponding current is finite if $z \neq 0$ and logarithmically infinite as z approaches 0. This is evident from (179) since $Y(\gamma)$ becomes inversely proportional to γ as γ increases indefinitely; and if $\exp(\gamma z) = 1$, the integral diverges as the series of reciprocals of integers. For a finite gap of length $s, T(\gamma) = V^i \sinh(\gamma s/2)/i\pi\gamma s$, provided the impressed field is assumed to be uniform; in this case, $I(z)$ is finite everywhere. When s is small, the field across the gap is substantially equal to the electrostatic field which would exist if the two halves of the infinite cylinder were kept at a given potential difference; thus a more accurate expression for $T(\gamma)$ can be obtained.

If the exciting field is produced by an ideal solenoid, or a magnetic loop carrying V^i volts, then

$$T(\gamma) = -V^i b\sqrt{-\gamma^2 - \beta^2}K_1(b\sqrt{-\gamma^2 - \beta^2})I_0(a\sqrt{-\gamma^2 - \beta^2})/2\pi i,\ b > a; \tag{180}$$

$$= V^i b\sqrt{-\gamma^2 - \beta^2}I_1(b\sqrt{-\gamma^2 - \beta^2})K_0(a\sqrt{-\gamma^2 - \beta^2})/2\pi i,\ b < a;$$

where b is the radius of the loop. If b approaches a in the external loop and if a approaches zero, then $T(\gamma)$ approaches $-V^i/2\pi i$.

To obtain all these formulas we should merely follow the pattern established in the preceding sections. The only really new feature is the occurrence of Bessel functions due to the cylindrical structure. To illustrate we shall derive (180). If $E_z(\rho,z)$ is the component of the electric intensity in the direction of the axis of the loop, then

$$E_z(b + 0,z) - E_z(b - 0,z) = V^i \int_C \frac{\sinh(\gamma s/2)}{i\pi\gamma s} e^{\gamma z}\, d\gamma, \tag{181}$$

where s becomes ultimately equal to zero. This equation is obtained from the electromagnetic laws when the loop is regarded as the limit of a magnetic current ribbon of radius b and width s. The electric intensity is continuous everywhere except across the ribbon; the total discontinuity in the axial voltage in passing through the ribbon is V^i. Next we express all the components of the field as Laplace integrals

$$E_z(\rho,z) = \int_C e_z(\rho,\gamma)e^{\gamma z}\, d\gamma, \qquad E_\rho(\rho,z) = \int_C e_\rho(\rho,\gamma)e^{\gamma z}\, d\gamma,$$
$$H_\varphi(\rho,z) = \int_C h_\varphi(\rho,\gamma)e^{\gamma z}\, d\gamma. \tag{182}$$

Since there is a discontinuity at $\rho = b$ which precludes a single set of analytic functions for *all* values of ρ and z, there will be a set of such

integrals for each region, $\rho < b$ and $\rho > b$. The connection between the two sets of functions is furnished by

$$e_z^+(b,\gamma) - e_z^-(b,\gamma) = V^i/2\pi i, \qquad h_\varphi^+(b,\gamma) = h_\varphi^-(b,\gamma), \tag{183}$$

where the superscript "plus" refers to the region $\rho > b$, and "minus" to the region $\rho < b$, and the width s of the ribbon has been assumed vanishingly small.

The longitudinal component of E satisfies the wave equation

$$\rho^2 \frac{\partial^2 E_z}{\partial \rho^2} + \rho \frac{\partial E_z}{\partial \rho} + \rho^2 \frac{\partial^2 E_z}{\partial z^2} + \beta^2 \rho^2 E_z = 0. \tag{184}$$

Substituting the Laplace integral and equating the integrand to zero, we have

$$\rho \frac{d^2 e_z}{d\rho^2} + \frac{de_z}{d\rho} + (\beta^2 + \gamma^2)\rho e_z = 0. \tag{185}$$

Expressing the solution in terms of the modified Bessel functions, we obtain

$$e_z^- = AI_0(\rho\sqrt{-\gamma^2 - \beta^2}), \qquad e_z^+ = BK_0(\rho\sqrt{-\gamma^2 - \beta^2}). \tag{186}$$

The first solution is the only one which is finite on the axis of the loop; the second is the only one which vanishes at infinity if the medium is dissipative, and which represents traveling waves moving *away* from the loop, if the medium is nondissipative. The corresponding transforms of the magnetic intensity are obtained from

$$h_\varphi = -\frac{i\omega\varepsilon}{\gamma^2 + \beta^2} \frac{\partial e_z}{\partial \rho}, \tag{187}$$

an equation which follows from Maxwell's equations. Differentiating, substituting in (183), and solving for A, B we find (180).

The proof of (179) is similar. Since the cylinder is assumed to be perfectly conducting, the total tangential E_z vanishes; hence the transform of E_z generated by currents on the cylinder should reduce to $-T(\gamma)$ for $\rho = a$. The same is true across the gap where, according to Newton's law of motion, the action should equal the reaction.

12. *Summary of the Laplace transform method and miscellaneous problems*

The various steps in the solution of problems by the Laplace transform method are:

1. Independent (given) functions in the differential equations may be such that they vanish outside some interval (x_1,x_2) of the independent

variable; if this is the case, calculate their Laplace transforms by (86). Otherwise, in order to insure the convergence of (86), set the given functions equal to zero for remote negative and positive values of x. Subsequently the interval of non-zero values may be extended as much as we wish.

2. If the given data involve discontinuities in the unknown functions, add to the given functions a set of impulse functions which would generate these discontinuities.
3. Introduce enough dependent (unknown) functions to recast the original differential equations into a system of first-order equations.
4. Express all functions as Laplace integrals and substitute in the equations. A set of equations is thus obtained for the Laplace transforms of the dependent functions. If the original set consists of ordinary differential equations, the new set will be a set of algebraic equations. If the original equations are partial differential equations containing two independent variables, the new equations will be ordinary differential equations. In any case, the Laplace transforms will satisfy equations containing $n - 1$ independent variables if the original equations contain n such variables.
5. Calculate the Laplace transforms of the unknown functions, taking into account, if necessary, the boundary conditions and the conditions at infinity.
6. Choose the proper contour of integration.
7. Evaluate the Laplace integrals representing the unknown functions.

The independent variable appearing in the Laplace integrals should not appear in the coefficients of the given differential equations; if it does, the method will not work (at step 4). If the given equations contain n independent variables of which m do not appear in the coefficients, the Laplace transform method will reduce the equations to another system, containing $n - m$ variables which do appear in the coefficients. The latter set will have to be solved by another method.

The independent variable appearing in the Laplace integrals should be capable of assuming all values in the range $(-\infty, \infty)$. This condition is satisfied automatically by the time t; but the coordinates in space are limited by the given system — as in the case of a semi-infinite line, extending from $x = 0$ to $x = \infty$, with a voltage source in series with the line at $x = \xi$, Figure 16.12. There are several ways in which we can get around this difficulty.

One is to use the *image method*. The line is extended indefinitely to the left and an image source is introduced at $x = -\xi$, Figure 16.12. If the given line is supposed to be open at $x = 0$, the current there must equal

zero and the image source must be equal and opposite to the source at $x = \xi$. On the other hand, if the given line is shorted the transverse voltage must equal zero, and the image source must be identical in strength and direction with the given source. The image source takes care of the boundary condition at $x = 0$ and the shorting wire can be removed, thereby rendering the line infinite and smooth. The new problem may be solved by the Laplace transform method; and for $x \geq 0$, the solution will coincide with the solution of the original problem.

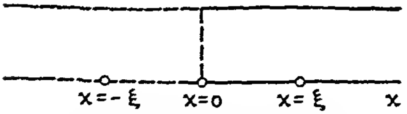

FIG. 16.12. Illustrating the image method of solving certain problems.

Another method of handling the situation is to extend the line to the left of $x = 0$ and solve the problem without assuming either image sources or discontinuities at $x = 0$. The solution will violate the given boundary conditions at $x = 0$. To set it straight we then assume a source of indefinite intensity at $x = 0$ and obtain the response to this source. Adding this solution to the previous one, we determine the intensity of the postulated source so as to satisfy the given boundary condition. That is, we equate to zero either the current or voltage at $x = 0$.

The two steps in the above method can be combined. To illustrate let us assume that the problem involves a short at $x = 0$ and a series impressed voltage V_0 at $x = \xi$. If the line is continued to the left of $x = 0$, then, on account of the short, $I(x)$ and $V(x)$ are identically zero for $x < 0$; the conditions for $x > 0$ remain unaffected. In both problems, the given and the modified, $V(0) = 0$. We shall make both problems identical for $x > 0$ if we assume, instead of the short, a sudden rise, $I(+0) - I(-0) = I_0$, in the current. For the time being, I_0 is left undetermined. The equations to be solved are

$$\frac{dV}{dx} = -i\omega LI + V_0 1(x,\xi),$$
$$\frac{dI}{dx} = -i\omega CV + I_0 1(x,0). \tag{188}$$

If $\hat{V}$ and $\hat{I}$ are the Laplace transforms of V and I,

$$\gamma\hat{V} = -i\omega L\hat{I} + \frac{V_0}{2\pi i}e^{-\gamma\xi}, \qquad \gamma\hat{I} = -i\omega C\hat{V} + \frac{I_0}{2\pi i}. \tag{189}$$

Solving,

$$\hat{V} = \frac{\gamma V_0 e^{-\gamma\xi} - i\omega LI_0}{2\pi i(\gamma^2 + \omega^2 LC)}, \qquad \hat{I} = \frac{\gamma I_0 - i\omega CV_0 e^{-\gamma\xi}}{2\pi i(\gamma^2 + \omega^2 LC)}. \tag{190}$$

When evaluating the Laplace integrals for $x < 0$, we can close the contour in the right half of the plane; thus we shall obtain

$$\left.\begin{aligned} V(x) &= \tfrac{1}{2}(-V_0 e^{-i\omega\sqrt{LC}\xi} + I_0\sqrt{L/C})e^{i\omega\sqrt{LC}x}, \\ I(x) &= \tfrac{1}{2}(-I_0 + V_0\sqrt{C/L}e^{-i\omega\sqrt{LC}\xi})e^{i\omega\sqrt{LC}x}. \end{aligned}\right\} \tag{191}$$

These expressions will vanish if

$$I_0 = V_0\sqrt{C/L}e^{-i\omega\sqrt{LC}\xi}. \tag{192}$$

Thus the unknown I_0 has been determined. It remains only to calculate $V(x)$ and $I(x)$ for $x > 0$. The two intervals, $(0,\xi)$ and (ξ,∞), have to be considered separately. In the interval $(0,\xi)$, neither right-handed nor left-handed closure is permissible for the complete integrands (190); the terms in the numerator have to be separated. The results are

$$\begin{aligned} V(x) &= -iV_0 e^{-i\omega\sqrt{LC}\xi}\sin(\omega\sqrt{LC}\,x), && 0 \le x < \xi; \\ &= V_0\cos(\omega\sqrt{LC}\,\xi)e^{-i\omega\sqrt{LC}x}, && \xi < x; \\ I(x) &= \sqrt{C/L}\,V_0 e^{-i\omega\sqrt{LC}\xi}\cos(\omega\sqrt{LC}\,x), && 0 \le x \le \xi; \\ &= \sqrt{C/L}\,V_0\cos(\omega\sqrt{LC}\,\xi)e^{-i\omega\sqrt{LC}x}, && \xi \le x. \end{aligned} \tag{193}$$

To review the execution of the various steps in the solution of problems by the Laplace transform method, let us consider an abstract problem:

$$\frac{d^2y}{du^2} = y + 3u, \qquad y'(0) = 0, \qquad y(0) = 2. \tag{194}$$

The independent function, $3u$, is not restricted; but let us change it to

$$\begin{aligned} f(u) &= 0, && u < 0; \\ &= 3u, && 0 < u < a; \\ &= 0, && u > a, \end{aligned} \tag{195}$$

in order to insure the convergence of the integrals. By (86) or (87) the Laplace transform, $\bar{f}(p)$, is

$$\bar{f}(p) = \frac{1}{2\pi i}\int_0^a 3ue^{-pu}\,du = \frac{3}{2\pi ip^2} - \frac{3}{2\pi ip^2}(1 + pa)e^{-pa}. \tag{196}$$

Note that $\bar{f}(p)$ consists of two parts, one of which is independent of a. Next we introduce another dependent function, $y_1 = dy/du$, and convert (194) into a system of first-order equations,

$$\frac{dy}{du} = y_1 + 2I(u,0), \qquad \frac{dy_1}{du} = y + f(u). \tag{197}$$

The impulse function will cause an upward discontinuity, and if the contour of integration is so chosen that $y(u)$ and $y_1(u)$ vanish for $u < 0$, the values of $y(+0)$ and $y_1(-0)$ will be as required by (194). We now express all functions as Laplace integrals

$$
\begin{aligned}
y(u) &= \int_C \hat{y}(p)e^{pu}\,dp, \qquad & y_1(u) &= \int_C \hat{y}_1(p)e^{pu}\,dp, \\
f(u) &= \int_C \hat{f}(p)e^{pu}\,dp, \qquad & I(u,0) &= \int T(p)e^{pu}\,dp,
\end{aligned}
\tag{198}
$$

where $T(p) = \lim\,[\sinh(ps/2)]/i\pi ps$, as $s \to 0$. The limit, $1/2\pi i$, is to be substituted only in the expressions for $\hat{y}(p)$ and $\hat{y}_1(p)$. If the unit impulse occurs when $u = \xi$, then

$$
I(u,\xi) = \int T(p)e^{pu}\,dp, \qquad T(p) = \lim \frac{\sinh(ps/2)}{i\pi s}\,e^{-p\xi} = \frac{e^{-p\xi}}{2\pi i}. \tag{199}
$$

Substituting from (198) in (197), differentiating where necessary, transferring all terms to one side, and equating the integrands to zero, we obtain

$$
p\hat{y} = \hat{y}_1 + (1/\pi i), \qquad p\hat{y}_1 = \hat{y} + \hat{f}. \tag{200}
$$

Note that these equations are obtained from (197) if we replace d/du by p and all the functions by their Laplace transforms. Solving (200), we have

$$
\hat{y}(p) = \frac{p}{i\pi(p^2 - 1)} + \frac{\hat{f}}{p^2 - 1}. \tag{201}
$$

Evaluating the Laplace integral for $y(u)$ in the interval $(0,a)$, we have

$$
y(u) = 2\cosh u + 3\sinh u - 3u. \tag{202}
$$

Note that in this interval $y(u)$ is independent of a, and if a is taken to be indefinitely large we have the solution of (194) and (197) in the interval $(0,\infty)$. Furthermore, (202) is the solution of (194) in $(-\infty,\infty)$ although in this extended interval it is not the solution of (197).

The formulation of differential equations and supplementary conditions depends on the physical laws governing the behavior of the system. Frequently the same set of equations expresses the behavior of different physical systems, because the particular physical aspect described by the equations is the same. For instance, the equations of motion of the mechanical system, Figure 16.13, possessing two degrees of freedom, are

$$
\begin{aligned}
M_1\frac{d^2x_1}{dt^2} + R_1\frac{dx_1}{dt} + S_1x_1 - S_2(x_2 - x_1) &= F_1(t), \\
S_2(x_2 - x_1) + M_2\frac{d^2x_2}{dt^2} + R_2\frac{dx_2}{dt} &= F_2(t).
\end{aligned}
\tag{203}
$$

The coordinates x_1, x_2 are the displacements of the masses M_1, M_2 from their neutral positions. The "two-mesh" electrical circuit in Figure 16.14 is also described by these equations, provided x_1, x_2 are interpreted as the quantities of displaced electric charge.

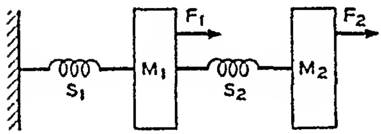

FIG. 16.13. A mechanical network obeying the equations of the electric network in Figure 16.14.

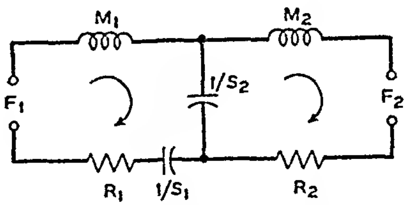

FIG. 16.14. An electric network obeying the equations of the mechanical network in Figure 16.13.

If $R_2 = F_2 = 0$, the second mesh will be resonant at some frequency. At this frequency, the second mesh will introduce an infinite impedance in the first mesh. If the frequency of F_1 coincides with this resonant frequency, the first mesh should, in the steady state, be at rest. It is suggested that the reader verify this result by solving (203). Mathematically, the result arises from the disappearance of a pole which normally yields the steady state term.

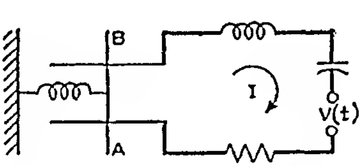

FIG. 16.15. An electromechanical network.

A two-mesh electromechanical circuit is shown in Figure 16.15, where the wire AB is supposed to oscillate in a steady magnetic field. The equations are

$$
\begin{aligned}
L\frac{dI}{dt} + RI + \frac{1}{C}\int^t I\,dt - Blv &= V(t), \\
BlI + M\frac{dv}{dt} + R_m v + S\int^t v\,dt &= F(t),
\end{aligned}
\tag{204}
$$

where v is the velocity of AB, l is the length of AB between the sliding contacts with the electrical circuit, and B is the density of the magnetic flux. The differences between (203) and (204) are superficial. The Laplace

transform method will reduce all these differential equations to algebraic equations.

The telegraphist's equations,*

$$\frac{\partial V}{\partial x} = -RI - L\frac{\partial I}{\partial t} + E(x,t),$$
$$\frac{\partial I}{\partial x} = -GV - C\frac{\partial V}{\partial t} - J(x,t), \tag{205}$$

contain two independent variables, x and t; two independent functions, $E(x,t)$ and $J(x,t)$; and two dependent functions, $V(x,t)$ and $I(x,t)$. Generally, the coefficients R, L, G, C may be functions of x and t; then if we try to use the Laplace transform method we shall find that it does not help us. If, however, the coefficients are independent of t, the Laplace transform method will reduce the partial differential equations to the ordinary equations,

$$\frac{d\hat{V}}{dx} = -R\hat{I} - Lp\hat{I} + \hat{E}(x,p),$$
$$\frac{d\hat{I}}{dx} = -G\hat{V} - Cp\hat{V} - \hat{J}(x,p). \tag{206}$$

Consider a semi-infinite line energized at $x = 0$; then $\hat{E} = \hat{J} = 0$. The conditions at $x = 0$ can be introduced after the solution has been obtained. Thus, if R, L, G, C are independent of x,

$$\hat{V}(x,p) = A(p)e^{-\Gamma x}, \qquad \hat{I}(x,p) = \frac{A(p)}{K(p)}e^{-\Gamma x},$$
$$\Gamma(p) = \sqrt{(R + pL)(G + pC)}\,, \qquad K(p) = \sqrt{\frac{R + pL}{G + pC}}, \tag{207}$$

where the real parts of Γ and K are non-negative. The arbitrary "constant of integration," $A(p)$, is a constant only as far as x is concerned. If a unit step voltage is impressed at $x = 0$,

$$\hat{V}(0,p) = A(p) = 1/2\pi ip, \qquad V(x,t) = \frac{1}{2\pi i}\int_C \frac{e^{pt-\Gamma x}}{p}\,dp. \tag{208}$$

If $R = G = 0$, the integral may be evaluated very simply. The expo-

* For applications of these equations to heat-flow problems, see: George H. Brown, Cyril N. Hoyler, and Rudolph A. Bierwirth, *Theory and Application of Radio-Frequency Heating*, D. Van Nostrand Company, Inc., New York, 1947, p. 146.

nential function becomes $\exp p(t - x\sqrt{LC})$. If $t < x\sqrt{LC}$, the right-handed closure is permissible and the integral vanishes. If $t > x\sqrt{LC}$, the left-handed closure is permissible and the integral is obtained from the residue at $p = 0$. The value is unity. The position of this voltage discontinuity is obtained from $t = x\sqrt{LC}$; thus the discontinuity is moving with the speed $dx/dt = 1/\sqrt{LC}$.

In the dissipative case we generally have two branch points

$$p_1 = -R/L, \qquad p_2 = -G/C. \tag{209}$$

In the exceptional case $R/L = G/C$, or $RC = GL$,

$$\Gamma = \sqrt{LC}\left(p + \frac{R}{L}\right) = p\sqrt{LC} + R\sqrt{C/L}, \tag{210}$$

the branch points disappear, and

$$\begin{aligned} V(x,t) &= \exp\left(-xR\sqrt{C/L}\right), \qquad & t > x\sqrt{LC}; \\ &= 0 & t < x\sqrt{LC}. \end{aligned} \tag{211}$$

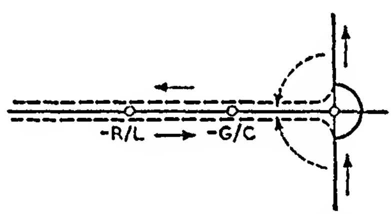

FIG. 16.16. A transformation of the normal contour of integration by swinging the imaginary semi-axes through 90°.

Regardless of the resistance R, the voltage discontinuity is moving with the speed $1/\sqrt{LC}$ and is attenuated with the distance from the source.

In the general case, the permissible closure of C is still determined by the sign of $t - x\sqrt{LC}$. For infinitely large values of p, R and G have no effect. Thus, in any case, the "wave front velocity" is $1/\sqrt{LC}$. If $t > x\sqrt{LC}$, we swing the contour as indicated in Figure 16.16 and obtain

$$\begin{aligned} V(x,t) = \exp\left(-x\sqrt{RG}\right) &+ \frac{1}{2\pi i}\int_{-R/L}^{-G/C} \frac{\exp\left[pt + ix\sqrt{-(R+pL)(G+pC)}\right]}{p}\,dp \\ &+ \frac{1}{2\pi i}\int_{-G/C}^{-R/L} \frac{\exp\left[pt - ix\sqrt{-(R+pL)(G+pC)}\right]}{p}\,dp. \end{aligned} \tag{212}$$

The first term is the contribution from the pole at the origin; it is the steady state term. The second term (the first integral) comes from the lower side of the real axis between $p = -R/L$ and $p = -G/C$, assuming that $R/L > G/C$. The last term is contributed by the upper side of the real axis. Along the remainder of the real axis, Γ has the same value, coming and going, and the corresponding integrals cancel.

Combining the integrals, we obtain

$$V(x,t) = \exp\left(-x\sqrt{RG}\right) + \frac{1}{\pi}\int_{-R/L}^{-G/C} \frac{e^{pt}}{p} \sin x\sqrt{-(R+pL)(G+pC)}\, dp. \tag{213}$$

The last term is the transient term since it decreases as t increases. It presents considerable computational difficulties.

For the further study of the Laplace transform method the reader is referred to treatises in various branches of applied mathematics.

CHAPTER XVII

The Gamma Function

1. *Definitions*

The coefficient of the mth derivative of x^n is $n(n-1)(n-2)\cdots(n-m+1)$. Similar products, in which the successive factors differ by unity, occur very frequently in the formation of power series. If n is an integer, such products can always be expressed in terms of special products of a certain number of integers beginning with unity. Thus, if we define the *factorial* of n,

$$n! = 1 \cdot 2 \cdot 3 \cdots n, \tag{1}$$

then

$$n(n-1)(n-2)\cdots(n-m+1) = n!/(n-m)!. \tag{2}$$

If n is not an integer, we can still keep equation (2) if we relax definition (1) and replace it by a difference equation

$$n! = (n-1)!n. \tag{3}$$

This definition would make $n! = (n-2)!(n-1)n$; and finally, $n! = (n-m)!(n-m+1)(n-m+2)\cdots n$.

Equation (3) does not possess a unique solution. Let $f(n)$ be another function satisfying

$$f(n) = nf(n-1). \tag{4}$$

Dividing these equations, we obtain

$$\frac{f(n)}{n!} = \frac{f(n-1)}{(n-1)!}. \tag{5}$$

Thus the ratio

$$\varphi(n) = \frac{f(n)}{n!} \tag{6}$$

is a periodic function, with the period unity. For instance, $\varphi(n)$ could equal $\sin 2\pi n$, $\exp(2i\pi n)$, $\cos 2\pi n$, $\tan \pi n$, $1 + \frac{1}{2}\cos 4\pi n$, etc. Any one solution of (4) would serve our purposes; but it is desirable to obtain a solution whose analytic behavior is not too complicated.

Euler proposed the following definition

$$z! = \int_0^\infty t^z e^{-t}\,dt = \int_0^\infty \exp(z \log t - t)\,dt. \tag{7}$$

The integral is convergent if re $(z) > -1$, that is, to the right of the straight line $x = -1$. Since

$$\begin{aligned}\int_0^\infty t^z e^{-t}\,dt &= -\int_0^\infty t^z\,de^{-t} = -t^z e^{-t}\Big|_0^\infty + z\int_0^\infty t^{z-1}e^{-t}\,dt \\ &= (z-1)!z,\end{aligned} \tag{8}$$

Euler's function satisfies (4); it reduces to (1) if z is a positive integer; when $z = 0$, we have

$$0! = 1. \tag{9}$$

Since Euler's function is defined everywhere in the strip $-1 < x \leq 0$, it can be extended to $x \leq -1$ by means of the reduction formula (3); thus

$$\begin{aligned}z! &= \frac{(z+1)!}{z+1} = \frac{(z+2)!}{(z+1)(z+2)} \\ &= \frac{(z+m)!}{(z+1)(z+2)\cdots(z+m)}.\end{aligned} \tag{10}$$

The extended function has simple poles at $z = -1, -2, -3, \cdots$.

Euler obtained another analytic expression for the factorial function

$$z! = \prod_1^\infty \left[\left(1+\frac{1}{n}\right)^z \left(1+\frac{z}{n}\right)^{-1}\right], \tag{11}$$

which is valid in the entire complex plane, except at the poles.*

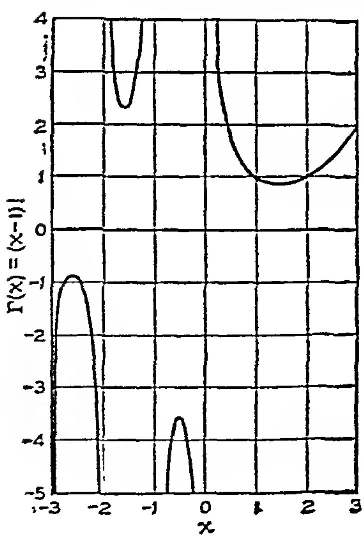

FIG. 17.1. The function $\Gamma(x) = (x-1)!$ for real values of x.

The symbol $\Pi(z)$ was introduced by Gauss to denote $z!$. Legendre introduced $\Gamma(z)$ to denote $(z-1)!$; thus

$$\Gamma(z) = (z-1)! = \Pi(z-1) \tag{12}$$

* Leonard Euler, Letter to Goldbach, Oct. 13, 1729. Corréspondence math. et phys. de quelques célèbres géomètres du 18^e siècle, publiée par Fuss, vol. I (St. Pétersburg, 1843).

is an identity. For real values of $z = x$, $\Gamma(x) = (x - 1)!$ and it varies as in Figure 17.1; the reciprocal varies as in Figure 17.2.

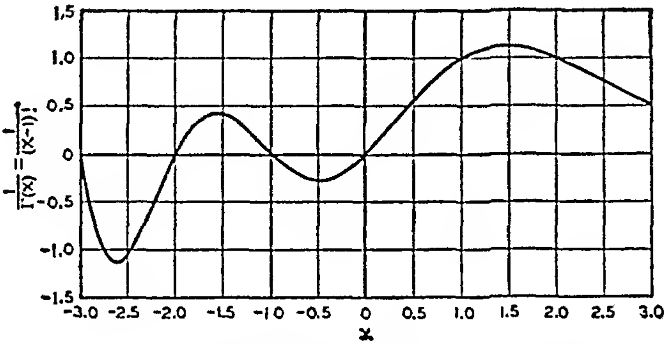

FIG. 17.2. The function $1/\Gamma(x) = 1/(x - 1)!$ for real values of x.

2. *Logarithmic derivatives*

The coefficients of the power series for Bessel functions, $J_n(x)$, Legendre functions, $P_n(x)$, and many other functions, may be expressed in terms of the factorials or gamma functions of the parameter n. The derivatives of gamma functions will thus occur in the series for $dJ_n(x)/dn$, $dP_n(x)/dn$, and other such functions. It is more convenient, however, to deal with the relative (logarithmic) derivatives

$$\Psi(z) = \frac{d}{dz} \log (z!) = \frac{1}{z!} \frac{d}{dz}(z!), \qquad \psi(z) = \frac{d}{dz} \log \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}. \tag{13}$$

In view of (12)

$$\Psi(z) = \psi(z + 1), \qquad \psi(z) = \Psi(z - 1). \tag{14}$$

Taking the logarithms of (10) and differentiating, we have

$$\begin{aligned} \Psi(z + m) &= \Psi(z) + \frac{1}{z + 1} + \frac{1}{z + 2} + \cdots + \frac{1}{z + m}, \\ \Psi(z + 1) &= \Psi(z) + \frac{1}{z + 1}, \qquad \Psi(z) = \Psi(z - 1) + \frac{1}{z}. \end{aligned} \tag{15}$$

The last two equations are the reduction formulas for the Ψ-function and correspond to equation (3) for the factorials.

3. *Weierstrass expansions*

Let us rewrite (15) as follows

$$\Psi(z+m) - \Psi(z) = \frac{1}{z+1} + \frac{1}{z+2} + \cdots + \frac{1}{z+m}. \tag{16}$$

As $m \to \infty$, the series does not converge. If, however, we subtract

$$\varphi(m) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m} \tag{17}$$

from (16), we obtain a convergent series

$$\lim\,[\Psi(z+m) - \Psi(z) - \varphi(m)] = \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right). \tag{18}$$

The typical term is $-z/n(n+z)$ which is dominated by k/n^2 where k is a suitable constant.

Let us take

$$\log\left(1+\frac{1}{n}\right) = \frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - \cdots, \tag{19}$$

choose successively $n = 1, 2, 3, \cdots m$, and add:

$$\log\,(m+1) = \varphi(m) - \sum_{1}^{m}\left(\frac{1}{2n^2} - \frac{1}{3n^3} + \cdots\right). \tag{20}$$

The limit of the sum exists and is called *Euler's constant*, C,

$$C = \lim\,[\varphi(m) - \log\,(m+1)] = \sum_{1}^{\infty}\left(\frac{1}{2n^2} - \frac{1}{3n^3} + \cdots\right), \tag{21}$$

$$C = 0.577215\cdots, \qquad e^C = 1.781072\cdots.$$

Hence, we can rewrite (18)

$$\lim\,[\Psi(z+m) - \Psi(z) - \log\,(m+1)] = C + \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right). \tag{22}$$

Integrating from $z = 0$ to $z = z$,

$$\begin{aligned}\lim\,[\log\,(z+m)! - \log m! - \log z! - z\log\,(m+1)]\\ = Cz + \sum_{n=1}^{\infty}\left[\log\left(1+\frac{z}{n}\right) - \frac{z}{n}\right].\end{aligned} \tag{23}$$

Taking the antilogarithms,

$$\lim \frac{(z+1)(z+2)\cdots(z+m)}{m!\,(m+1)^z} = e^{Cz}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)e^{-z/n}. \tag{24}$$

The left side is the reciprocal of the first m terms of Euler's infinite product (11); hence

$$z! = e^{-Cz} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{z/n}. \tag{25}$$

This expansion was given by Weierstrass.

The corresponding expansions for the logarithmic derivative $\Psi(z)$, and for $\Psi'(z)$ are

$$\Psi(z) = -C - \sum_{n=1}^{\infty} \left(\frac{1}{z+n} - \frac{1}{n}\right), \qquad \Psi'(z) = \sum_{n=1}^{\infty} \frac{1}{(z+n)^2}. \tag{26}$$

If we take a resistance $R_n = 1/n$ in parallel with an inductance $L_n = 1/n^2$, and connect the circuits corresponding to $n = 1, 2, 3, \cdots$ in series, then the impedance of the network will be $\Psi(p) + C$, where p is the oscillation constant. The impedance at the frequency $f = \omega/2\pi$ will thus be $\Psi(i\omega) + C$.

Similarly, take an inductance $L_n = 1$, resistance $R_n = 2n$, and capacitance $C_n = 1/n^2$, and connect them in series; connect such circuits for $n = 1, 2, 3, \cdots$ in parallel; then the admittance of the network will be $p\Psi'(p)$.

Further information about factorials and gamma functions may be obtained from E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, and from a paper by J. L. W. V. Jensen, "An Elementary Exposition of the Theory of Gamma Functions," *Annals of Mathematics*, Second Series, Vol. 17, No. 3, March, 1916.

4. *Formulas for reference*

The following are some of the most useful formulas connecting factorials and related functions.

$$x! = x(x-1)!, \qquad 0! = \Gamma(1) = 1, \qquad 1! = 1. \tag{27}$$

$$(-\tfrac{1}{2})! = \Gamma(\tfrac{1}{2}) = \sqrt{\pi}. \tag{28}$$

$$(-x)!\,(x-1)! = \pi \csc \pi x = \Gamma(x)\Gamma(1-x). \tag{29}$$

$$(-x)! = \pi/(x-1)! \sin \pi x. \tag{30}$$

$$\left(-\frac{1}{n}\right)!\left(-\frac{2}{n}\right)!\left(-\frac{3}{n}\right)!\cdots\left(-\frac{n-1}{n}\right)! = (2\pi)^{(n-1)/2} n^{-1/2}. \tag{31}$$

$$x!\,(x-\tfrac{1}{2})! = \pi^{1/2}\, 2^{-2x} (2x)!. \tag{32}$$

$$\left(\frac{x}{2}\right)!\left(\frac{x-1}{2}\right)! = \pi^{1/2} 2^{-x}\, x!. \tag{33}$$

$$x!\left(x-\frac{1}{n}\right)!\left(x-\frac{2}{n}\right)!\cdots\left(x-\frac{n-1}{n}\right)! = (2\pi)^{(n-1)/2}n^{-1/2}n^{-nx}(nx)!. \quad (34)$$

$$x! \sim \sqrt{2\pi x}\, x^x e^{-x}\left(1+\frac{1}{12x}+\frac{1}{288x^2}-\cdots\right). \quad (35)$$

$$2^n(n-\tfrac{1}{2})! = 1\cdot 3\cdot 5\cdots(2n-1)\sqrt{\pi}. \quad (36)$$

$$\Psi(x) = \Psi(x-1)+x^{-1}. \quad (37)$$

$$\Psi(-x) = \Psi(x-1)+\pi\cot\pi x. \quad (38)$$

$$\Psi(x-\tfrac{1}{2}) = \Psi(-x-\tfrac{1}{2})+\pi\tan\pi x. \quad (39)$$

$$\Psi(0) = -C = -0.577215\cdots; \qquad \Psi(-n) = \infty. \quad (40)$$

$$\Psi(n) = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-C. \quad (41)$$

$$\Psi(-\tfrac{1}{2}) = -C-2\log 2 = -1.96351\cdots. \quad (42)$$

$$\Psi\left(-\frac{1}{2}\pm n\right) = \Psi\left(-\frac{1}{2}\right)+2\left(1+\frac{1}{3}+\frac{1}{5}+\cdots+\frac{1}{2n-1}\right). \quad (43)$$

$$\Psi(0.4616\cdots) = 0. \quad (44)$$

$$\Psi'(x) = \Psi'(x-1)-x^{-2}. \quad (45)$$

$$-\Psi'(-x) = \Psi'(x-1)-\pi^2\csc^2\pi x. \quad (46)$$

$$-\Psi'(-x-\tfrac{1}{2}) = \Psi'(x-\tfrac{1}{2})-\pi^2\sec^2\pi x. \quad (47)$$

CHAPTER XVIII

Exponential Integrals

1. *Definitions*

The following group of integrals is important in the theory of radiation and other fields of applied mathematics:

$$
\begin{aligned}
\operatorname{Ein} z &= \int_0^z \frac{1 - e^{-w}}{w}\, dw, \qquad z = x + iy, \\
\operatorname{Ein}(iy) &= \operatorname{Cin} y + i \operatorname{Si} y, \\
\operatorname{Cin} y &= \int_0^y \frac{1 - \cos t}{t}\, dt, \qquad \operatorname{Si} y = \int_0^y \frac{\sin t}{t}\, dt, \\
\operatorname{Ci} y &= \int_\infty^y \frac{\cos t}{t}\, dt = C + \log y - \operatorname{Cin} y,
\end{aligned}
\tag{1}
$$

where C is Euler's constant, defined by (17–21). The first two functions are called *exponential integrals;* the remaining functions are *cosine* and *sine integrals.* Ein z, Cin z and Si z are analytic except at infinity; Ci z has a logarithmic singularity at the origin.

2. *Power series*

Expanding the exponential function and integrating term by term,

$$\operatorname{Ein} z = \sum_{n=1}^{\infty} (-)^{n-1} \frac{z^n}{n!\,n}. \tag{2}$$

Substituting $z = iy$ and separating the real and imaginary parts,

$$
\begin{aligned}
\operatorname{Cin} y &= \frac{y^2}{2!2} - \frac{y^4}{4!4} + \frac{y^6}{6!6} - \cdots, \\
\operatorname{Si} y &= y - \frac{y^3}{3!3} + \frac{y^5}{5!5} - \cdots.
\end{aligned}
\tag{3}
$$

3. *Asymptotic series*

For large values of the variable the following asymptotic series are useful

(see Problem 7, Section 15.9, for the value of Si ∞):

$$\int_x^\infty t^{-1}e^{-t}\,dt \sim x^{-1}e^{-x}(1 - x^{-1} + 2!x^{-2} - 3!x^{-3} + \cdots),$$

$$\operatorname{Si} y \sim \frac{\pi}{2} - \frac{\cos y}{y}\sum_{n=0}^{\infty}\frac{(-)^n(2n)!}{y^{2n}} - \frac{\sin y}{y}\sum_{n=0}^{\infty}\frac{(-)^n(2n+1)!}{y^{2n+1}}, \tag{4}$$

$$\operatorname{Ci} y \sim \frac{\sin y}{y}\sum_{n=0}^{\infty}\frac{(-)^n(2n)!}{y^{2n}} - \frac{\cos y}{y}\sum_{n=0}^{\infty}\frac{(-)^n(2n+1)!}{y^{2n+1}}.$$

To derive these series we integrate by parts as follows:

$$\begin{aligned}\int_x^\infty t^{-1}e^{-t}\,dt &= -\int_x^\infty t^{-1}\,de^{-t} = -t^{-1}e^{-t}\Big|_x^\infty - \int_x^\infty t^{-2}e^{-t}\,dt\\ &= x^{-1}e^{-x} + \int_x^\infty t^{-2}\,de^{-t}\\ &= x^{-1}e^{-x} - x^{-2}e^{-x} - \int_x^\infty t^{-3}e^{-t}\,dt.\end{aligned}$$

The derivation of the asymptotic series for Si y and Ci y is left to the reader.

4. *Asymptotic formulas for* Ein z

From (1) we have

$$\begin{aligned}\operatorname{Ein} z &= \int_0^1 \frac{1-e^{-w}}{w}\,dw + \int_1^z \frac{1-e^{-w}}{w}\,dw\\ &= \int_0^1 \frac{1-e^{-w}}{w}\,dw + \log z - \int_1^z \frac{e^{-w}}{w}\,dw\\ &= A + \log z + \int_z^\infty \frac{e^{-w}}{w}\,dw,\end{aligned}\tag{5}$$

where

$$A = \int_0^1 \frac{1-e^{-w}}{w}\,dw - \int_1^\infty \frac{e^{-w}}{w}\,dw. \tag{6}$$

If the real part of z is non-negative, the last integral in (5) approaches zero as z becomes infinite.

The evaluation of A depends on a certain amount of foresight, on a feeling that the integrals in (6) are related to Euler's constant C. The logarithmic function is obviously involved on account of dw/w; the exponential terms remind one of the integral (17–7) for the factorial, although the latter diverges for $z = -1$. The integrals in (6) however are convergent.

All this indicates that we might come nearer to our goal if we convert C, as given by (17–21), into an integral. First we note that

$$\frac{1}{m} = \int_0^1 t^{m-1}\,dt, \qquad \log n = \int_1^n \frac{dt}{t},$$

$$\begin{aligned} 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} &= \int_0^1 (1+t+t^2+\cdots+t^{n-1})\,dt \\ = \int_0^1 \frac{1-t^n}{1-t}\,dt &= \int_0^1 \frac{1-(1-u)^n}{u}\,du = \int_0^n t^{-1}\left[1-\left(1-\frac{t}{n}\right)^n\right]dt; \end{aligned} \tag{7}$$

therefore,

$$\begin{aligned} C &= \lim\left[\int_0^n \left[1-\left(1-\frac{t}{n}\right)^n\right]\frac{dt}{t} - \int_1^n \frac{dt}{t}\right], \quad n \to \infty, \\ &= \lim\left[\int_0^1 \left[1-\left(1-\frac{t}{n}\right)^n\right]\frac{dt}{t} - \int_1^n \left(1-\frac{t}{n}\right)^n \frac{dt}{t}\right]. \end{aligned} \tag{8}$$

We also have

$$\begin{aligned} \left(1-\frac{t}{n}\right)^n &= \exp\left[n\log\left(1-\frac{t}{n}\right)\right] = \exp\left[-n\left(\frac{t}{n}+\frac{t^2}{2n^2}+\cdots\right)\right] \\ &= \exp\left(-t-\frac{t^2}{2n}-\frac{t^3}{3n^2}-\cdots\right) \\ &= e^{-t}\exp\left(-\frac{t^2}{2n}-\frac{t^3}{3n^2}-\cdots\right) \\ &= e^{-t}\left(1-\frac{t^2}{2n}-\frac{t^3}{3n^2}-\cdots\right), \end{aligned} \tag{9}$$

if n is greater than t. Substituting in (8) and passing to the limit, we obtain

$$C = \int_0^1 \frac{1-e^{-t}}{t}\,dt - \int_1^\infty \frac{e^{-t}}{t}\,dt. \tag{10}$$

It must be emphasized, however, that it is not enough to establish that the expression in parentheses in (8) approaches $\exp(-t)$; the difference between the expression and its limit must be small enough for the integral of this difference to vanish in the limit. Suppose the difference had been constant and equal to $1/n$; then for any value of n, no matter how large, the integral of the error would have been infinite. It can be shown, however, that in the present case it is permissible to pass to the limit.

Thus the constant (6) is equal to Euler's constant and (5) becomes

$$\operatorname{Ein} z = C + \log z + \int_z^\infty \frac{e^{-w}}{w}\,dw. \tag{11}$$

If $z = iy$, we obtain the fourth equation in (1).

Problems

1. Show that

$$\int_a^b x^{-1} e^{-(\alpha+i\beta)x}\,dx = (\operatorname{Ci}\beta b - \operatorname{Ci}\beta a) - i(\operatorname{Si}\beta b - \operatorname{Si}\beta a)$$

$$-2(\alpha/\beta)\sin\tfrac{1}{2}\beta(b-a)\exp\left[-\tfrac{1}{2}i\beta(b+a)\right]$$

$$+\sum_{n=2}^{\infty}(-)^n\frac{\alpha^n}{n!}\int_a^b x^{n-1}e^{-i\beta x}\,dx.$$

Note that the series is convergent for all values of α.

2. Show that

$$\int_a^b x^{-1}e^{-(\alpha+i\beta)x}\,dx = \int_{\alpha a}^{\alpha b} x^{-1}e^{-x}\,dx + \sum_{n=1}^{\infty}\frac{(-i\beta)^n}{n!\,\alpha^n}\int_{\alpha a}^{\alpha b} x^{n-1}e^{-x}\,dx.$$

5. *Derivatives and integrals*

The derivatives of the exponential integrals follow from their definitions

$$\frac{d}{dx}\operatorname{Ein} x = \frac{1-e^{-x}}{x}, \qquad \frac{d}{dx}\operatorname{Cin} x = \frac{1-\cos x}{x},$$

$$\frac{d}{dx}\operatorname{Ci} x = \frac{\cos x}{x}, \qquad \frac{d}{dx}\operatorname{Si} x = \frac{\sin x}{x}. \tag{12}$$

The integrals are obtained if we integrate by parts,

$$\int_0^x \operatorname{Ein} t\,dt = t\operatorname{Ein} t\Big|_0^x - \int_0^x (1-e^{-t})\,dt$$

$$= 1 - x - e^{-x} + x\operatorname{Ein} x. \tag{13}$$

Similarly

$$\int_0^x \operatorname{Cin} t\,dt = x\operatorname{Cin} x - x + \sin x,$$

$$\int_0^x \operatorname{Ci} t\,dt = x\operatorname{Ci} x - \sin x, \tag{14}$$

$$\int_0^x \operatorname{Si} t\,dt = x\operatorname{Si} x + \cos x - 1.$$

The integrals of the above integrals can also be expressed in terms of the exponential integrals and elementary functions.

Problems

1. Show that

$$\int_0^x \frac{\sin^2 t}{t^2}\,dt = \frac{\cos 2x - 1}{2x} + \operatorname{Si} 2x,$$

$$\int_x^\infty \frac{\sin t}{t^2}\,dt = \frac{\sin x}{x} - \operatorname{Ci} x, \qquad \int_x^\infty \frac{\cos t}{t^2}\,dt = -\frac{\pi}{2} + \operatorname{Si} x + \frac{\cos x}{x}.$$

2. Show that

$$\int_0^\pi \frac{\cos^2(\frac{1}{2}\pi\cos\theta)}{\sin\theta}\,d\theta = \tfrac{1}{2}\operatorname{Cin} 2\pi.$$

3. Show that

$$\int_0^\pi \frac{1-\cos[\beta\ell(1-\cos\theta)]}{(1-\cos\theta)^2}\sin^3\theta\,d\theta = 2\left(\operatorname{Cin} 2\beta\ell - 1 + \frac{\sin 2\beta\ell}{2\beta\ell}\right).$$

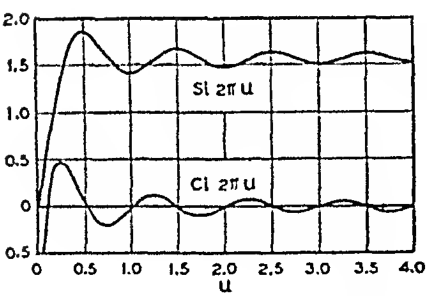

FIG. 18.1. The functions Si $2\pi u$ and Ci $2\pi u$.

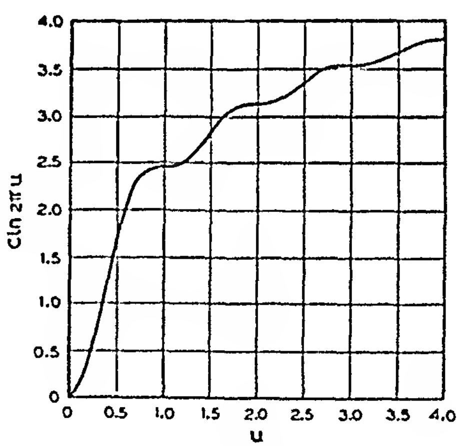

FIG. 18.2. The function Cin $2\pi u$.

6. *Curves and tables*

Figures 18.1 and 18.2 illustrate the behavior of Ci $2\pi u$, Si $2\pi u$, Cin $2\pi u$. Frequently the Ci and Si functions occur jointly, as in Ci $2\pi u \pm i$ Si $2\pi u$; in such cases the spiral representations, Figures 18.3 and 18.4, are convenient.

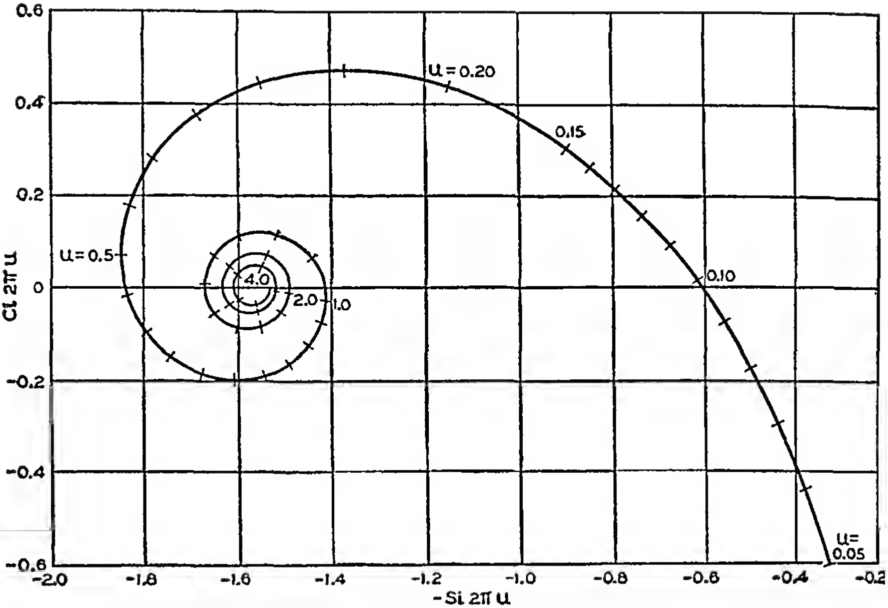

FIG. 18.3. A spiral representing Ci $2\pi u$ vs. $-$Si $2\pi u$.

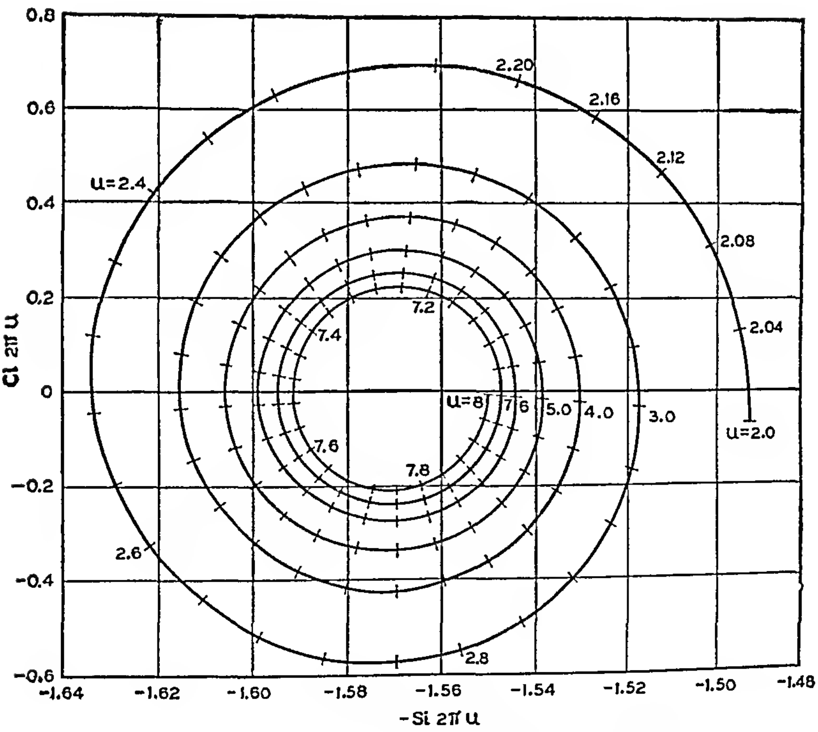

FIG. 18.4. A spiral representing Ci $2\pi u$ vs. $-$Si $2\pi u$.

While the sine and cosine integrals are fluctuating functions, the functions

$$
\begin{aligned}
A &= \text{am}\,[\text{Ci}\,2\pi u + i\,\text{Si}\,2\pi u - i\pi/2], \\
\Phi &= \text{ph}\,[\text{Ci}\,2\pi u + i\,\text{Si}\,2\pi u - i\pi/2] - 360°\,u - 180°,
\end{aligned}
\tag{15}
$$

are monotonic. In terms of A and Φ we have

$$
\begin{aligned}
\text{Ci}\,2\pi u &= -A\cos\,(\Phi + 360°\,u), \\
\text{Si}\,2\pi u &= \tfrac{1}{2}\pi - A\sin\,(\Phi + 360°\,u).
\end{aligned}
\tag{16}
$$

A table of A and Φ is given below.

Table I

u	A	Φ degrees	u	A	Φ
0	∞	0	0.46	0.312402	74.4095
0.01	2.659842	30.9373	.48	.301131	74.8763
.02	2.083592	36.7182	.50	.290632	75.3168
.03	1.767065	40.6871	.55	.267285	76.3168
.04	1.554013	43.7724	.60	.247354	77.1937
.05	1.396330	46.3145	.65	.230144	77.9686
.06	1.272910	48.4824	.70	.215136	78.6582
.07	1.172646	50.3740	.75	.201935	79.2758
.08	1.088996	52.0517	.80	.190235	79.8319
.09	1.017792	53.5582	.85	.179796	80.3351
.10	.956221	54.9240	.90	.170426	80.7926
.11	.902302	56.1722	.95	.161972	81.2102
.12	.854590	57.3201	1.00	.154303	81.5926
.13	.811997	58.3815	1.1	.140930	82.2687
.14	.773690	59.3676	1.2	.129665	82.8470
.15	.739014	60.2873	1.3	.120048	83.3471
.16	.707447	61.1481	1.4	.111746	83.7835
.17	.678569	61.9652	1.5	.104507	84.1674
.18	.652032	62.7170	1.6	.098140	84.5077
.19	.627550	63.4351	1.7	.092499	84.8113
.20	.604884	64.1142	1.8	.087466	85.0847
.22	.564217	65.3692	1.9	.082949	85.3294
.24	.528740	66.5044	2.0	.078873	85.5522
.26	.497494	67.5375	2.2	.071807	85.9402
.28	.469746	68.4826	2.4	.065899	86.2671
.30	.444929	69.3511	2.6	.060884	86.5457
.32	.422594	70.1524	2.8	.056576	86.7861
.34	.402382	70.8944	3.0	.052835	86.9954
.36	.383999	71.5836	3.2	.049557	87.1799
.38	.367207	72.2257	3.4	.046660	87.3424
.40	.351805	72.8255	3.6	.044083	87.4876
.42	.337628	73.3870	3.8	.041775	87.6179
.44	.324533	73.9140	4.0	.039696	87.7354

The maxima and minima of Si x occur when $\sin x$ vanishes; likewise the maxima and minima of Ci x occur when $\cos x$ vanishes. The curve for Cin x has points of inflection at $x = 2n\pi$. A brief table of these functions for $x = n\pi/2$ is given below.

Table II

x	$\pi/2$	π	$3\pi/2$	2π	$5\pi/2$	3π
Cin x	0.55680	1.64827	2.32584	2.43765	2.51446	2.80993
Ci x	0.47200	0.07367	−0.19844	−0.02256	0.12377	0.01063
Si x	1.37076	1.85194	1.60837	1.41815	1.55583	1.67476

CHAPTER XIX

FRESNEL INTEGRALS

1. *Definitions*

Fresnel integrals are defined as

$$C(x) = \int_0^x \cos(\pi t^2/2)\,dt, \qquad S(x) = \int_0^x \sin(\pi t^2/2)\,dt. \tag{1}$$

They occur in the theory of diffraction of waves. Figure 19.1 illustrates the behavior of these functions.

Let

$$u = C(x), \qquad v = S(x) \tag{2}$$

be the parametric equations of a curve (u is the abscissa and v is the ordi-

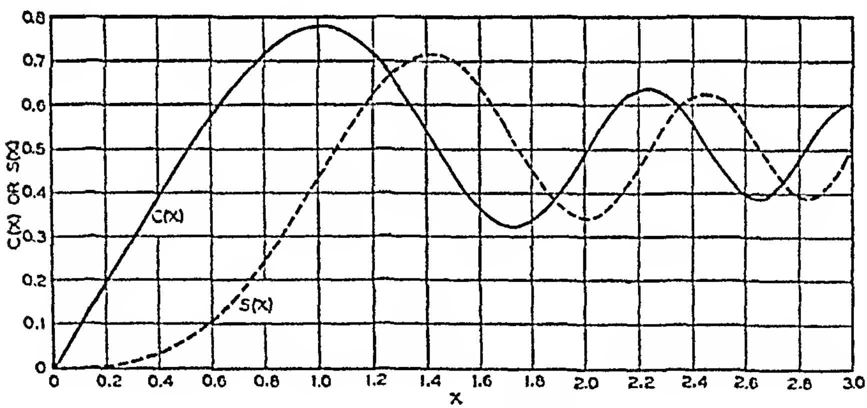

FIG. 19.1. The Fresnel integrals.

nate); this curve, Figure 19.2, is called *Cornu's spiral*. The length along the curve, from the origin to some point (u,v), is the independent variable x,

$$s = \int_0^x \sqrt{du^2 + dv^2} = \int_0^x dx = x. \tag{3}$$

Fresnel integrals are odd functions,

$$C(-x) = -C(x), \qquad S(-x) = -S(x); \tag{4}$$

hence Cornu's spiral is symmetric about the origin. In diffraction theory, Cornu's spiral permits a rapid appraisal of variations in the wave intensity in a given region.

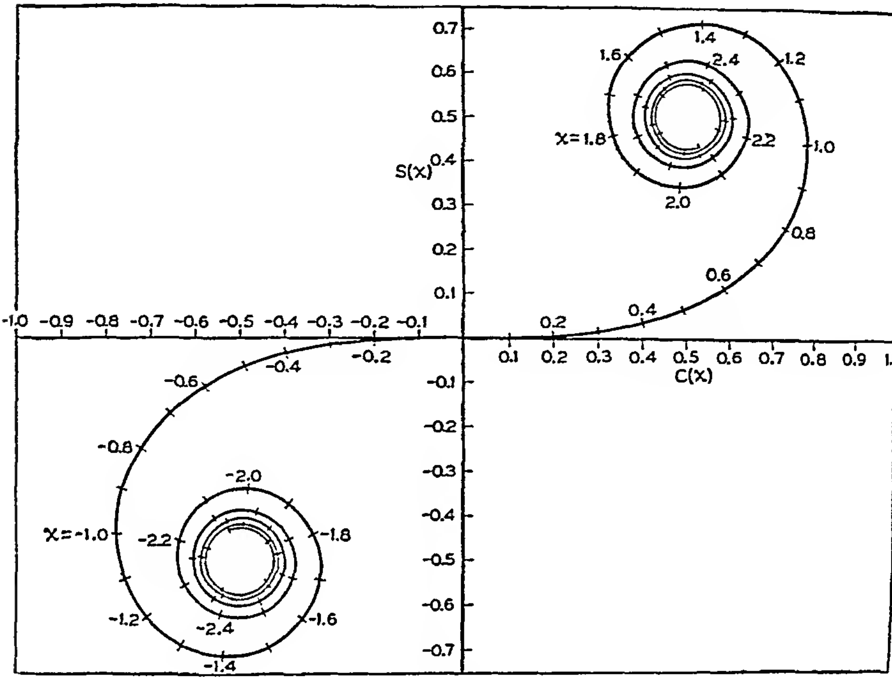

FIG. 19.2. The Cornu spiral.

2. *Power series*

Expanding the integrands in (1) in power series and integrating term by term, we have

$$C(x) = \sum_{n=0}^{\infty} \frac{(-)^n \pi^{2n} x^{4n+1}}{2^{2n}(2n)!(4n+1)},$$

$$S(x) = \sum_{n=0}^{\infty} \frac{(-)^n \pi^{2n+1} x^{4n+3}}{2^{2n+1}(2n+1)!(4n+3)}. \tag{5}$$

3. *Asymptotic expansions*

For large values of x, we have [see Section 15.9 for the values of $C(\infty)$ and $S(\infty)$]·

$$
\begin{aligned}
&C(x) \sim \tfrac{1}{2} + P(x)\cos(\pi x^2/2) - Q(x)\sin(\pi x^2/2),\\
&S(x) \sim \tfrac{1}{2} + P(x)\sin(\pi x^2/2) + Q(x)\cos(\pi x^2/2),\\
&P(x) = \frac{1}{\pi x}\sum_{n=0}^{\infty}\frac{(-)^{n+1}1\cdot 3\cdot 5\cdots(4n+1)}{(\pi x^2)^{2n+1}},\\
&Q(x) = \frac{1}{\pi x}\sum_{n=0}^{\infty}\frac{(-)^{n+1}1\cdot 3\cdot 5\cdots(4n-1)}{(\pi x^2)^{2n}}.
\end{aligned}
\tag{6}
$$

The first term of $Q(x)$ is $-1/\pi x$.

These expansions are derived as follows

$$
\begin{aligned}
C(x) + iS(x) &= \int_0^x \exp(i\pi t^2/2)\,dt\\
&= C(\infty) + iS(\infty) + \int_\infty^x \exp(i\pi t^2/2)\,dt,\\
\int_\infty^x \exp(i\pi t^2/2)\,dt &= \frac{1}{i\pi}\int_\infty^x t^{-1} d\exp(i\pi t^2/2)\\
&= \frac{\exp(i\pi x^2/2)}{i\pi x} + \frac{1}{i\pi}\int_\infty^x t^{-2}\exp(i\pi t^2/2)\,dt.
\end{aligned}
\tag{7}
$$

Integration by parts is continued so as to obtain a series of decreasing powers of x. The factorials in the numerators of P and Q in (6) make the series divergent for all values of x; but if only n terms are taken, the remainder is represented by an integral tending to zero as $1/x^{2n-1}$ if x increases indefinitely.

Problems

1. Show that

$$
\begin{aligned}
&\int_{-b/2}^{b/2}\int_{-a/2}^{a/2} \exp\frac{i\pi}{\lambda z}[(x-u)^2 + (y-v)^2]\,du\,dv\\
&= \tfrac{1}{2}\lambda z\left[C\left(\frac{a+2x}{\sqrt{2\lambda z}}\right) + C\left(\frac{a-2x}{\sqrt{2\lambda z}}\right) - iS\left(\frac{a+2x}{\sqrt{2\lambda z}}\right) - iS\left(\frac{a-2x}{\sqrt{2\lambda z}}\right)\right]\\
&\times\left[C\left(\frac{b+2y}{\sqrt{2\lambda z}}\right) + C\left(\frac{b-2y}{\sqrt{2\lambda z}}\right) - iS\left(\frac{b+2y}{\sqrt{2\lambda z}}\right) - iS\left(\frac{b-2y}{\sqrt{2\lambda z}}\right)\right].
\end{aligned}
$$

2. Evaluate

$$\left| \int_{-a/2}^{a/2} \cos\,(\pi\hat{x}/a) e^{i\beta\hat{z}}\, dx \right|^2, \qquad \hat{z} = \frac{a^2}{8\ell} - \frac{\hat{x}^2}{2\ell}, \qquad \beta = \frac{2\pi}{\lambda}.$$

Ans. $\frac{1}{2}\ell\lambda\,\{[C(u) - C(v)]^2 + [S(u) - S(v)]^2\}$, where

$$u = \frac{1}{\sqrt{2}}\left(\frac{\sqrt{\lambda\ell}}{a} + \frac{a}{\sqrt{\lambda\ell}}\right), \qquad v = \frac{1}{\sqrt{2}}\left(\frac{\sqrt{\lambda\ell}}{a} - \frac{a}{\sqrt{\lambda\ell}}\right).$$

CHAPTER XX

Bessel Functions

1. *Bessel's equation and standard forms of its solutions*

From a purely mathematical point of view Bessel's equation of order ν,

$$x^2 \frac{d^2y}{dx^2} + x \frac{dy}{dx} + (x^2 - \nu^2) y = 0, \tag{1}$$

is just a linear differential equation with nothing very remarkable to distinguish it from scores of other similar equations. Its claim to a distinctive name is based on its importance in physical applications, particularly in wave theory. The solutions of (1) are called *Bessel functions;* their order ν may be either real or complex. The general solution could be expressed in terms of any two linearly independent solutions; but for various reasons more than two have been denoted by distinctive symbols.

Figures 20.1 and 20.2 illustrate the behavior of two kinds of Bessel functions for real values of x and small positive values of ν. Both kinds look alike for larger values of x, when they are very nearly sinusoidal functions with amplitudes varying inversely as the square root of x. In the vicinity of the origin the functions differ: Bessel functions of the *first kind* are finite when $x = 0$, those of the *second kind* are infinite. These properties could have been anticipated on physical grounds. Bessel functions occur in the theory of cylindrical and spherical waves, just as sinusoidal functions appear in the theory of plane waves. The independent variable x is proportional to the distance from the line or point sources. When x is small the curvature of a wave front is important and the functions differ from simple sine functions; but when x is large, the waves begin to resemble plane waves and Bessel functions resemble sine functions. Presently we shall draw these qualitative conclusions from purely numerical analysis of the equation. This analogy will be exploited in the " sinusoidal " interpolation and extrapolation of Bessel functions.

Several sections in this chapter are devoted to the calculation of Bessel functions. Tables are available only for certain values of ν and one should be able to supplement them. The ascending power series for Bessel functions converge for all values of x; but such series are never very suitable when x is large. Thus it becomes necessary to subdivide the range of the independent variable into three parts: (1) small values of x, (2)

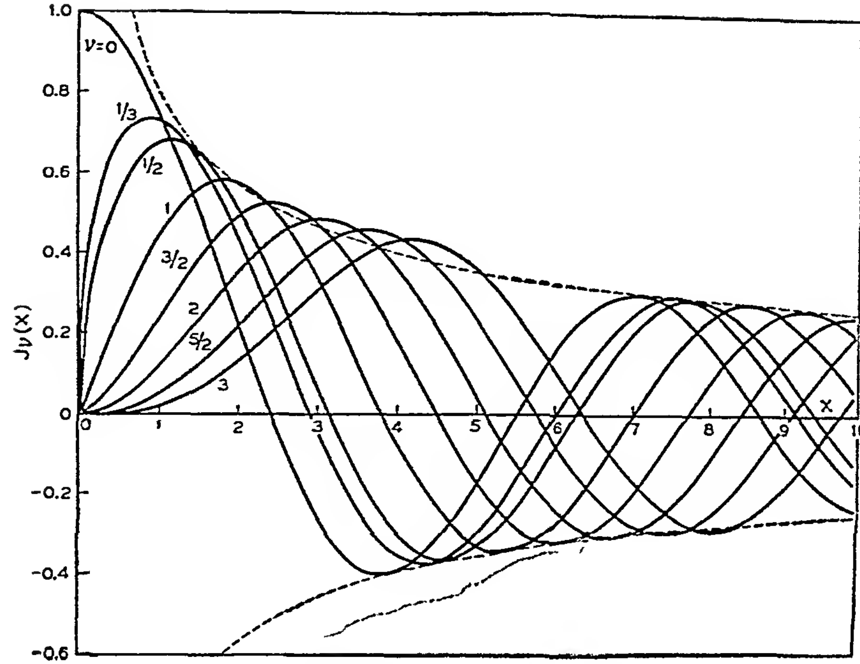

Fig. 20.1. Bessel functions of the first kind of order ν.

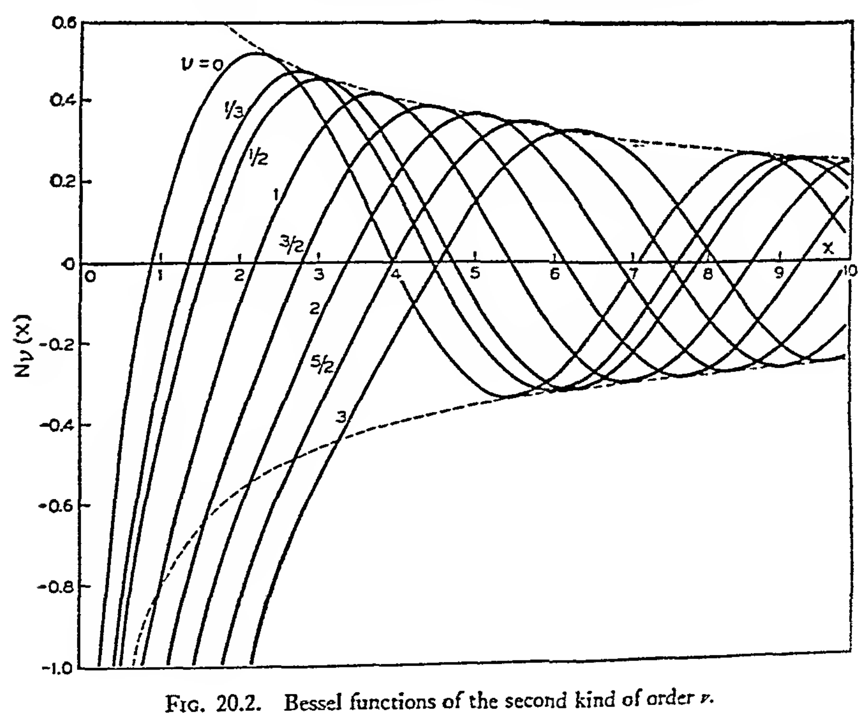

Fig. 20.2. Bessel functions of the second kind of order ν.

intermediate values of x, and (3) large values of x. Bessel functions possess properties which enable us at times to make substantial analytical simplifications; the more important of these are considered in later sections.

The power series for Bessel functions have already been derived in Section 11.6

$$\begin{aligned} J_\nu(x) &= \frac{x^\nu}{\nu!\,2^\nu}\left[1 - \frac{(x/2)^2}{\nu+1} + \frac{(x/2)^4}{1\cdot 2(\nu+1)(\nu+2)} - \cdots\right], \\ J_{-\nu}(x) &= \frac{2^\nu x^{-\nu}}{(-\nu)!}\left[1 - \frac{(x/2)^2}{-\nu+1} + \frac{(x/2)^4}{1\cdot 2(-\nu+1)(-\nu+2)} - \cdots\right]. \end{aligned} \tag{2}$$

The numerical factors $\nu!\,2^\nu$ and $(-\nu)!\,2^{-\nu}$ are included arbitrarily. The factorials of negative numbers may be eliminated by using (17–30); thus, if ν is positive and exceeds an integer n by a proper fraction,

$$\begin{aligned} J_{-\nu}(x) = {} & \frac{(x/2)^{-\nu}\sin\nu\pi}{\pi}\sum_{m=0}^{n-1}\frac{(\nu-m-1)!(x/2)^{2m}}{m!} \\ & + \sum_{m=n}^{\infty}\frac{(-)^m(x/2)^{-\nu+2m}}{m!(-\nu+m)!}. \end{aligned} \tag{3}$$

If $\nu = n$,

$$J_{-n}(x) = (-)^n J_n(x), \tag{4}$$

and the two solutions are no longer linearly independent. To form the second solution we might start with

$$\frac{J_\nu(x) - (-)^n\, J_{-\nu}(x)}{\nu - n}, \tag{5}$$

and find its limit as ν approaches n; this is in fact how the solution of the " second kind " was originally constructed by Hankel. There are definite advantages, however, to a definition which makes the solution an analytic function of ν as well as of x. Weber and Schläfli proposed the following function

$$N_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi}. \tag{6}$$

Another common symbol for this function is $Y_\nu(x)$. As ν approaches n, $\cos\nu\pi$ approaches $(-)^n$; the numerator and denominator tend to zero but the limit of their ratio exists

$$N_n(x) = \lim N_\nu(x) = \frac{1}{\pi}\left[\frac{\partial J_\nu}{\partial\nu} + (-)^{n+1}\frac{\partial J_{-\nu}}{\partial\nu}\right]. \tag{7}$$

Presently we shall carry out the differentiation and obtain a series for $N_n(x)$.

Any solution of the Bessel equation may be expressed in terms of the above solutions; in wave theory, however, certain other forms are useful. If the Bessel equation had been encountered in wave theory in the first instance, it is very probable that these other forms would have been introduced as standard solutions. *Bessel's equation describes a two-dimensional wave*, with x denoting the distance from the center of the disturbance. For large x, the wave front is nearly straight and it is difficult to distinguish the wave from a one-dimensional wave in which the disturbance is proportional to $\exp(-ix)$. The power carried by the wave per unit length of the wave front is proportional to the square of the amplitude A and the total power is proportional to the product of A^2 and the circumference $2\pi x$ of the wave front. The law of conservation of energy requires that this power be independent of x so that A should be proportional to $1/\sqrt{x}$; hence we should expect a solution of the Bessel equation which is proportional to $x^{-1/2}\exp(-ix)$ for large x.

To verify this conjecture we divide (1) by x^2,

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{\nu^2}{x^2}\right)y = 0. \tag{8}$$

As x increases indefinitely, the equation seems to approach an equation with constant coefficients whose solutions are $\exp(\mp ix)$ or $\cos x$ and $\sin x$. There is, however, an uncertainty about the cumulative effect of the second term. Let us remove this uncertainty (see Section 11.7). Introducing a new dependent variable

$$= w/\sqrt{x} \tag{9}$$

in (8), we have

$$\frac{d^2w}{dx^2} = -w + \frac{\nu^2 - 0.25}{x^2}\,w. \tag{10}$$

If this equation is treated by the wave perturbation method, it is found that for large x the first term is a sinusoidal function and the second term approaches zero as x increases. Hence for large x,

$$y = \frac{Ae^{ix} + Be^{-ix}}{\sqrt{x}} = \frac{C\cos x + D\sin x}{\sqrt{x}}. \tag{11}$$

It so happens that the following combinations of J_ν and N_ν,

$$H_\nu^{(1)}(x) = J_\nu(x) + iN_\nu(x), \qquad H_\nu^{(2)}(x) = J_\nu(x) - iN_\nu(x), \tag{12}$$

behave as exponential functions of gradually diminishing amplitude. In fact,

$$
\begin{aligned}
H_\nu^{(1)}(x) &\sim (2/\pi x)^{1/2} \exp i\,(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi), \\
H_\nu^{(2)}(x) &\sim (2/\pi x)^{1/2} \exp i\,(-x + \tfrac{1}{2}\nu\pi + \tfrac{1}{4}\pi); \\
J_\nu(x) &\sim (2/\pi x)^{1/2} \cos\,(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi), \\
N_\nu(x) &\sim (2/\pi x)^{1/2} \sin\,(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi).
\end{aligned}
\tag{13}
$$

The H functions are called *Hankel's* functions. They are analogous to exponential functions just as J_ν and N_ν are analogous to sinusoidal functions. Figures 20.1 and 20.2 represent J_ν and N_ν for smaller values of ν; the dotted lines represent the amplitude factor $(2/\pi x)^{1/2}$.

If $\nu = \frac{1}{2}$, the last term in (10) vanishes and the corresponding expressions (13) are exact for all x; thus

$$
\begin{aligned}
J_{1/2}(x) &= (2/\pi x)^{1/2} \sin x, & N_{1/2}(x) &= -(2/\pi x)^{1/2} \cos x; \\
H_{1/2}^{(1)}(x) &= -i(2/\pi x)^{1/2}\, e^{ix}, & H_{1/2}^{(2)} &= i(2/\pi x)^{1/2}\, e^{-ix}.
\end{aligned}
\tag{14}
$$

Later in this chapter it is shown that if $\nu = n + \frac{1}{2}$, Bessel functions can be expressed in closed forms involving circular functions and polynomials in $1/x$. In this case

$$N_{n+1/2}(x) = (-)^{n+1} J_{-(n+1/2)}(x). \tag{15}$$

These functions play an important role in problems involving spherical symmetry. In view of this the name, "cylinder functions," sometimes applied to Bessel functions, is not quite appropriate.

2. *Modified Bessel's equation*

The modified Bessel's equation,

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + \nu^2)y = 0, \tag{16}$$

is obtained from Bessel's equation when x is replaced by ix. The solutions are called the *modified Bessel functions*. For large real x these functions behave as exp $(\mp x)$, cosh x, and sinh x, all divided by the square root of x.

The two most important linearly independent solutions are

$$I_\nu(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{\nu+2m}}{m!(\nu + m)!}, \tag{17}$$

$$K_\nu(x) = \tfrac{1}{2}\pi \csc \nu\pi\,[I_{-\nu}(x) - I_\nu(x)]. \tag{18}$$

For positive values of ν, the I functions are finite at $x = 0$ and the K func-

tions are not; when the real part of x is non-negative, the K functions vanish at $x = \infty$ and the I functions do not. Under certain circumstances these properties eliminate either I or K as a possible solution of a physical problem.

For large x the K function is represented asymptotically by

$$K_\nu(x) \sim (\pi/2x)^{1/2} e^{-x}, \tag{19}$$

provided $-3\pi/2 < \text{ph}\,(x) < 3\pi/2$. Similarly, if $-\pi/2 < \text{ph}\,(x) < 3\pi/2$, then

$$I_\nu(x) \sim (2\pi x)^{-1/2}[e^x + e^{-x+i\nu\pi+i\pi/2}]. \tag{20}$$

Restrictions on the phase are necessary because, in general, the modified Bessel functions and their asymptotic representations are multiple-valued functions whose branches do not correspond uniquely.

If $\nu = \frac{1}{2}$, then

$$K_{1/2}(x) = (\pi/2x)^{1/2}e^{-x}, \qquad I_{1/2}(x) = (2/\pi x)^{1/2}\sinh x, \tag{21}$$

for all x. More generally, if $\nu = n + \frac{1}{2}$, where n is an integer, the I and K functions are expressible in closed form in terms of exponential functions and polynomials in $1/x$.

The modified Bessel functions are particularly convenient for treatment of waves in dissipative media, whereas Bessel functions are preferable in problems involving nondissipative media.

3. *Differential equations reducible to Bessel's equation*

A number of simple equations may be reduced to Bessel's equation or the modified Bessel's equation by simple transformations of the independent and dependent variables. Thus denoting by $B(x)$ the general solution of Bessel's equation, we have the following table:

$$\begin{aligned}
&\frac{d^2y}{dx^2} - \frac{2\nu - 1}{x}\frac{dy}{dx} + y = 0, && y = x^\nu B_\nu(x);\\
&\frac{d^2y}{dx^2} = -xy, && y = x^{1/2}B_{1/3}(\tfrac{2}{3}x^{3/2});\\
&\frac{d^2y}{dx^2} = -x^{-1/2}y, && y = x^{1/2}B_{2/3}(\tfrac{4}{3}x^{3/4});\\
&\frac{d^2y}{dx^2} = -\mu^2\nu^2x^{2\mu-2}y, && y = x^{1/2}B_{1/2\mu}(\nu x^\mu);\\
&\frac{d^2y}{dx^2} = -y + \frac{\nu^2 - 0.25}{x^2}y, && y = x^{1/2}B_\nu(x);
\end{aligned} \tag{22}$$

$$x\frac{d^2y}{dx^2}+(1-\nu)\frac{dy}{dx}+\frac{1}{4}y=0, \qquad y=x^{\nu/2}B_\nu(\sqrt{x});$$

$$\frac{d^2y}{dx^2}+\frac{1}{x}\frac{dy}{dx}+\frac{1}{4}\left(\frac{1}{x}-\frac{\nu^2}{x^2}\right)y=0, \qquad y=B_\nu(\sqrt{x}); \tag{22}$$

$$\frac{d^2y}{dx^2}+\frac{1}{x}\frac{dy}{dx}+4\left(x^2-\frac{\nu^2}{x^2}\right)y=0, \qquad y=B_\nu(x^2).$$

4. *Infinite series for $N_n(x)$ and $K_n(x)$*

To obtain the series for $N_n(x)$, we differentiate the power series (2) and (3) with respect to ν and substitute in (7). A typical term of J_ν is

$$\frac{(-)^m(x/2)^{\nu+2m}}{m!(\nu+m)!}=\frac{(-)^m\exp[(\nu+2m)\log(x/2)]}{m!(\nu+m)!}; \tag{23}$$

its derivative with respect to ν is

$$\frac{(-)^m(x/2)^{\nu+2m}}{m!(\nu+m)!}\log(x/2)-\frac{(-)^m(x/2)^{\nu+2m}}{m!}\frac{\frac{d}{d\nu}[(\nu+m)!]}{[(\nu+m)!]^2}$$
$$=\frac{(-)^m(x/2)^{\nu+2m}}{m!(\nu+m)!}\log(x/2)-\frac{(-)^m(x/2)^{\nu+2m}}{m!(\nu+m)!}\Psi(\nu+m), \tag{24}$$

where Ψ is the logarithmic derivative of the factorial. Thus

$$\left.\frac{\partial J_\nu}{\partial\nu}\right|_{\nu=n}=J_n(x)\log(x/2)-\sum_{m=0}^{\infty}\frac{(-)^m(x/2)^{n+2m}}{m!(n+m)!}\Psi(n+m),$$
$$\Psi(n+m)=-C+1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n+m}=-C+\varphi(n+m), \tag{25}$$

where C is Euler's constant. This constant is associated with all terms of the power series for J_n and therefore

$$\left.\frac{\partial J_\nu}{\partial\nu}\right|_{\nu=n}=J_n(x)(\log x-\log 2+C)-\sum_{m=0}^{\infty}\frac{(-)^m(x/2)^{n+2m}}{m!(n+m)!}\varphi(n+m). \tag{26}$$

Similarly the derivative of the infinite series in (3) is

$$\sum_{m=n}^{\infty}\frac{(-)^m(x/2)^{2m-n}}{m!(m-n)!}[-\log(x/2)+\Psi(m-n)]$$
$$=(-)^{n+1}\sum_{m=0}^{\infty}\frac{(-)^m(x/2)^{n+2m}}{m!(n+m)!}[\log(x/2)-\Psi(m)] \tag{27}$$
$$=(-)^{n+1}\left[J_n(x)(\log x-\log 2+C)-\sum_{m=1}^{\infty}\frac{(-)^m(x/2)^{n+2m}}{m!(n+m)!}\varphi(m)\right].$$

The limit of the derivative of the finite series in (3) as $\nu \to n$ may be obtained simply by differentiating $\sin \nu\pi$ and replacing ν by n; thus we have

$$(-)^n \sum_{m=0}^{n-1} \frac{(n-m-1)!(x/2)^{-n+2m}}{m!}. \tag{28}$$

Combining these results and substituting in (7), we have

$$\begin{aligned} N_n(x) = &-\frac{1}{\pi} \sum_{m=0}^{n-1} \frac{(n-m-1)!(x/2)^{-n+2m}}{m!} \\ &+ \frac{2}{\pi} J_n(x)\,(\log x - \log 2 + C) \\ &- \frac{1}{\pi} \sum_{m=0}^{\infty} \frac{(-)^m (x/2)^{n+2m}}{m!(n+m)!}\,[\varphi(m) + \varphi(n+m)]. \end{aligned} \tag{29}$$

If $n = 0$, this form is not directly applicable. Retracing the above derivation for this special case, we obtain

$$N_0(x) = \frac{2}{\pi} J_0(x)\,(\log x + C - \log 2) - \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-)^m (x/2)^{2m}}{(m!)^2}\,\varphi(m). \tag{30}$$

Similarly, we can obtain the series for K_n,

$$\begin{aligned} K_n(x) = &\frac{1}{2} \sum_{m=0}^{n-1} \frac{(-)^m (n-m-1)!(x/2)^{-n+2m}}{m!} \\ &+ (-)^{n+1}\,(\log x + C - \log 2) I_n(x) \\ &+ (-)^n \frac{1}{2} \sum_{m=0}^{\infty} \frac{(x/2)^{n+2m}}{m!(n+m)!}\,[\varphi(m) + \varphi(n+m)], \end{aligned} \tag{31}$$

$$K_0(x) = -(\log x + C - \log 2) I_0(x) + \sum_{m=1}^{\infty} \frac{(x/2)^{2m}}{(m!)^2}\,\varphi(m).$$

All these series converge for real and complex values of x, except when $x = 0$.

5. *Bessel functions of order $\nu = n + 1/2$ and asymptotic series*

If $\nu = n + 1/2$, where n is an integer, then Bessel functions and modified Bessel functions may be expressed as finite series of circular and exponential functions multiplied by negative powers of the independent variable. Presently we shall prove that

$$\begin{aligned} K_{n+1/2}(x) &= (\pi/2x)^{1/2} \hat{K}_n(x), \qquad I_{n+1/2}(x) = (2/\pi x)^{1/2} \hat{I}_n(x), \\ J_{n+1/2}(x) &= (2/\pi x)^{1/2} \hat{J}_n(x), \qquad N_{n+1/2}(x) = (2/\pi x)^{1/2} \hat{N}_n(x), \end{aligned} \tag{32}$$

where

$$
\begin{aligned}
&\hat{K}_0(x) = e^{-x}, \qquad \hat{K}_1(x) = e^{-x}\left(1 + \frac{1}{x}\right), \\
&\hat{K}_2(x) = e^{-x}\left(1 + \frac{3}{x} + \frac{3}{x^2}\right), \\
&\hat{I}_0(x) = \sinh x, \qquad \hat{I}_1(x) = \cosh x - \frac{\sinh x}{x}, \\
&\hat{I}_2(x) = \left(1 + \frac{3}{x^2}\right)\sinh x - \frac{3}{x}\cosh x, \\
&\hat{J}_0(x) = \sin x, \qquad \hat{J}_1(x) = \frac{\sin x}{x} - \cos x, \\
&\hat{J}_2(x) = \left(\frac{3}{x^2} - 1\right)\sin x - \frac{3}{x}\cos x, \\
&\hat{N}_0(x) = -\cos x, \qquad \hat{N}_1(x) = -\sin x - \frac{\cos x}{x}, \\
&\hat{N}_2(x) = \left(1 - \frac{3}{x^2}\right)\cos x - \frac{3}{x}\sin x,
\end{aligned}
\tag{33}
$$

and, in general,

$$
\begin{aligned}
&\hat{K}_n(x) = e^{-x}\sum_{m=0}^{n}\frac{(n+m)!}{m!(n-m)!(2x)^m}, \\
&\hat{I}_n(x) = \tfrac{1}{2}[\hat{K}_n(-x) + (-)^{n+1}\hat{K}_n(x)], \\
&\hat{J}_n(x) + i\hat{N}_n(x) = (-i)^{n+1}\hat{K}_n(-ix), \\
&\hat{J}_n(x) = [\cos(n\pi/2)\sin x - \sin(n\pi/2)\cos x]\,A_n(x) \\
&\qquad\qquad + [\cos(n\pi/2)\cos x + \sin(n\pi/2)\sin x]\,B_n(x), \\
&\hat{N}_n(x) = [\cos(n\pi/2)\sin x - \sin(n\pi/2)\cos x]\,B_n(x) \\
&\qquad\qquad - [\cos(n\pi/2)\cos x + \sin(n\pi/2)\sin x]\,A_n(x), \\
&A_n(x) = \sum_{m=0}^{2m\le n}\frac{(-)^m(n+2m)!}{(2m)!(n-2m)!(2x)^{2m}}, \\
&B_n(x) = \sum_{m=0}^{2m\le n-1}\frac{(-)^m(n+2m+1)!}{(2m+1)!(n-2m-1)!(2x)^{2m+1}}.
\end{aligned}
\tag{34}
$$

The $\hat{K}_n$ and $\hat{I}_n$ functions are solutions of

$$\frac{d^2w}{dx^2} = w + \frac{n(n+1)}{x^2}\,w. \tag{35}$$

Similarly, $\hat{J}_n$ and $\hat{N}_n$ satisfy

$$\frac{d^2w}{dx^2} = -w + \frac{n(n+1)}{x^2}\,w. \tag{36}$$

These equations are obtained from the corresponding equations for Bessel functions by making a substitution which removes the terms containing the first derivative and setting $\nu = n + 1/2$.

In order to prove the first equation in (34) we let

$$w = e^{-x}\left(1 + \frac{a_1}{x} + \frac{a_2}{x^2} + \frac{a_3}{x^3} + \cdots\right) \tag{37}$$

and substitute in (35). The first term in parentheses is chosen to be in accord with the asymptotic behavior of K_ν as given by (19). Performing the necessary differentiation and comparing the coefficients, we obtain the following recurrence formulas

$$a_m = \frac{n(n+1) - m(m-1)}{2m}\,a_{m-1} = \frac{(n+m)(n+1-m)}{2m}\,a_{m-1}. \tag{38}$$

Hence

$$a_1 = \tfrac{1}{2}n(n+1), \qquad a_2 = \tfrac{1}{8}(n-1)n(n+1)(n+2), \cdots. \tag{39}$$

The series terminates with $m = n$.

If n is not an integer, the series does not terminate. *Formally* the series still satisfies the differential equation; but it diverges for all values of x. Nevertheless, the series remains useful for large values of x, for it is an asymptotic series. If only a fixed number of terms of the series are retained, they will represent $K_{\nu+1/2}$ with increasing accuracy as x increases.

6. *Approximations for large values of the independent variable*

Applying the wave perturbation method to (10), we obtain

$$\begin{aligned} H_\nu^{(1)}(x) &\cong (2/\pi x)^{1/2} \exp \frac{-i(2\nu+1)\pi}{4} \\ &\quad \times \left\{e^{ix} - (\nu^2 - \tfrac{1}{4})\left[\operatorname{Ci} 2x + i\left(\operatorname{Si} 2x - \frac{\pi}{2}\right)\right] e^{-ix}\right\}, \\ H_\nu^{(2)}(x) &\simeq (2/\pi x)^{1/2} \exp \frac{i(2\nu+1)\pi}{4} \\ &\quad \times \left\{e^{-ix} - (\nu^2 - \tfrac{1}{4})\left[\operatorname{Ci} 2x - i\left(\operatorname{Si} 2x - \frac{\pi}{2}\right)\right] e^{ix}\right\} \end{aligned} \tag{40}$$

as far as the first perturbation. The second perturbation has to be evaluated numerically. As $\nu^2 - 0.25$ approaches zero, the approximation becomes more accurate even for small values of x. When $\nu = 0.5$, the formula is exact.

The approximation (40) deteriorates as ν increases, unless x also increases, and at a faster rate. For large x the amplitude of the perturbation term varies inversely as x; thus x should increase roughly as ν^2 if this term is to remain of the same order of magnitude relative to the first term.

To obtain an approximation of Liouville's type we substitute

$$q^2 = \nu^2 - \tfrac{1}{4}, \qquad q = \sqrt{\nu^2 - \tfrac{1}{4}}, \qquad \nu \geq \tfrac{1}{2}, \tag{41}$$

in (10) and apply (11–65); thus

$$w(x) = \frac{Ae^{\pm i\Phi}}{\sqrt[4]{1 - (q/x)^2}}, \qquad \Phi = \int^x \sqrt{1 - (q/x)^2}\, dx. \tag{42}$$

To integrate Φ, let

$$x = q \sec \beta, \qquad dx = q \sec \beta \tan \beta \, d\beta; \tag{43}$$

thus

$$\Phi = q \int^{\beta} \tan^2 \beta \, d\beta = q\,(\tan \beta - \beta). \tag{44}$$

The constant of integration is omitted because it can always be absorbed in the arbitrary constant A. Next we substitute in (42), then in (9), and determine A in such a way that as x and β approach infinity, $y(x)$ approaches either one or the other Hankel function whose asymptotic behavior is defined in (13); thus we find

$$\begin{gathered} J_\nu(x) \pm i N_\nu(x) \sim \frac{\cos \vartheta \pm i \sin \vartheta}{\sqrt{\tfrac{1}{2}\pi q \tan \beta}}, \\ \vartheta = q\,(\tan \beta - \beta) - \tfrac{1}{4}\pi + \tfrac{1}{2}(q - \nu)\pi. \end{gathered} \tag{45}$$

From (41) we obtain

$$q = \nu\left(1 - \frac{1}{4\nu^2}\right)^{1/2} = \nu - \frac{1}{8\nu} - \frac{1}{128\nu^3} - \frac{1}{1024\nu^5} - \cdots. \tag{46}$$

These formulas tend to become exact as q approaches zero.

Similar approximations are obtained from (11–66), (11–67), and (45) if we replace q by ν and let $x = \nu \sec \beta$. These approximations are the ones usually found in the literature; they are Liouville's approximations applied directly to Bessel's equation.

In terms of x we have

$$
\begin{aligned}
q \tan \beta &= \sqrt{x^2 - q^2}, \\
q\,(\tan \beta - \beta) &= \sqrt{x^2 - q^2} - q \tan^{-1} \sqrt{(x/q)^2 - 1}.
\end{aligned} \tag{47}
$$

Liouville's approximation for K_ν is

$$K_\nu(x) = \frac{\sqrt{\pi/2}}{\sqrt[4]{(x^2 + q^2)}} \exp\left(-\sqrt{x^2 + q^2} + q \sinh^{-1} \frac{q}{x}\right). \tag{48}$$

This also becomes exact for all x when $q = 0$, $\nu = 1/2$.

Using the result of Problem 5, Section 11.11, we find that the exact function $v = [1 - (q/x)^2]^{1/4} w(x)$ satisfies

$$
\begin{aligned}
\frac{d^2 v}{d\Phi^2} &= -\left[1 + \frac{5q^4}{4(x^2 - q^2)^3} + \frac{3q^2}{2(x^2 - q^2)^2}\right] v \\
&= -\left[1 + \frac{5}{4q^2 \tan^6 \beta} + \frac{3}{2q^2 \tan^4 \beta}\right] v.
\end{aligned} \tag{49}
$$

From this we conclude that for a fixed value of β Liouville's approximation improves as q increases and that it is a good approximation only if $x - q$ is substantially larger than the cube root of q (at least two or three times as large).

The zeros of J_ν and N_ν are approximately equal to the zeros of $\cos \vartheta$ and $\sin \vartheta$; thus as q increases we have asymptotically

$$x \sim q + 0.5 \sqrt[3]{9\left[\vartheta + \frac{\pi}{4} + \frac{1}{2}(\nu - q)\pi\right]^2}\, \sqrt[3]{q}, \tag{50}$$

$$x \sim \nu + 0.5 \sqrt[3]{9\left(\vartheta + \frac{\pi}{4}\right)^2}\, \sqrt[3]{\nu},$$

where $\vartheta = \frac{1}{2}\pi + m\pi$ for the zeros of J_ν and $\vartheta = m\pi$ (including $m = 0$) for the zeros of N_ν. The following table compares the exact (Jahnke and Emde) and approximate values of the coefficient of the cube root of ν for the successive zeros of J_ν

	1	2	3
exact	1.8557...	3.2447...	4.3817...
approx.	1.8415...	3.2397...	4.37898...

For the first zero of N_ν the approximate value of the coefficient is 0.885 while the exact value is $0.931577 \cdots$; this is the smallest root and the

approximation is not particularly good. The approximate values of the next two roots are 2.59 and 3.83; the error should be substantially less than one per cent.

7. *Approximations for small values of the independent variable*

For small values of x, power series furnish good approximations. The series (2) for J_ν converges rapidly if $x < \sqrt{\nu + 1}$. Similarly the series for $J_{-\nu}$ converges fairly fast if $x < \sqrt{\nu - 1}$. In the case of N_ν, where $\nu = n + \delta$ and $0 < \delta < 1$, we obtain from (3) and (6)

$$N_\nu(x) = J_\nu(x)\cot \nu\pi - \frac{2^\nu(\nu-1)!}{\pi x^\nu}\left[1 + \frac{(x/2)^2}{\nu - 1} + \frac{(x/2)^4}{2(\nu-1)(\nu-2)} + \cdots \right.$$
$$\left. + \frac{(x/2)^{2n-2}}{(n-1)!(\nu-1)(\nu-2)\cdots(\delta+1)}\right] \tag{51}$$
$$- \frac{(x/2)^{n-\delta}}{n!(-\delta)!\sin\delta\pi}\left[1 - \frac{(x/2)^2}{(n+1)(1-\delta)} + \cdots\right].$$

When ν is an integer we find from (29)

$$N_n(x) = -\frac{(n-1)!2^n}{\pi x^n}\left[1 + \frac{(x/2)^2}{n-1} + \frac{(x/2)^4}{2(n-1)(n-2)} + \cdots\right],$$
$$N_0(x) = \frac{2}{\pi}(\log x + C - \log 2)\,J_0(x) + \frac{x^2}{2\pi} - \cdots. \tag{52}$$

Approximations of Liouville's type are better for an extended range of x. In the preceding section we applied Liouville's approximation to equation (10); but when x is small, we find from (49) that for small values of q the perturbation term is large. For this reason we shall apply the approximation directly to Bessel's equation as in Section 11.11. If in (11–67) we change n to ν and write

$$x = \nu\,\mathrm{sech}\,\alpha, \qquad u = \int^x \left(\frac{\nu^2}{x^2} - 1\right)^{1/2} dx = -\nu(\alpha - \tanh\alpha), \tag{53}$$

we have

$$y(x) = \frac{A\exp\left[\pm\nu(\alpha - \tanh\alpha)\right]}{\sqrt{\nu\tanh\alpha}}. \tag{54}$$

The constant may be determined to match the behavior of these approximations with that of the standard forms of Bessel functions as x approaches

zero. Thus

$$J_\nu(x) \sim \frac{x^\nu \exp\left(\sqrt{\nu^2 - x^2} - \nu\right)}{[1 + \sqrt{1 - (x/\nu)^2}]^\nu \nu! \sqrt[4]{1 - (x/\nu)^2}},$$
$$N_\nu(x) \sim -\frac{x^{-\nu}(\nu - 1)!\,[1 + \sqrt{1 - (x/\nu)^2}]^\nu \exp\left(\nu - \sqrt{\nu^2 - x^2}\right)}{\pi\sqrt[4]{1 - (x/\nu)^2}}. \tag{55}$$

There is a weakness in the identification of the second approximation with N_ν. As x approaches zero, $N_\nu + BJ_\nu$ will approach the same asymptotic expression, regardless of the value of the constant B. Naturally it would have a negligible effect for sufficiently small values of x; but it will restrict the upper limit for which the approximation is still satisfactory. However, the same expression may be obtained by another method which is free from this objection. It should also be noted that the objection applies only to the identification of the approximation with a particular solution of the second kind; a linear combination of two Liouville's approximations will give a perfectly general approximation to the complete solution of Bessel's equation.

To estimate the range in which Liouville's approximation is satisfactory, we should carry out the substitution of the new independent variable into Bessel's equation and then free the new equation from the first derivative by changing the dependent variable. Thus we find that the exact equation for $v(x) = y(x)\sqrt[4]{\nu^2 - x^2}$ is

$$\frac{d^2v}{du^2} = \left[1 - \frac{x^2(x^2 + 4\nu^2)}{4(\nu^2 - x^2)^3}\right]v. \tag{56}$$

The second term in the brackets should be small compared with unity. As ν increases the range increases; and then the condition is that $\nu - x$ should be several times as large as $1/\sqrt[3]{\nu}$.

8. *Approximations in the intermediate region*

Approximations of Liouville's type break down in the vicinity of $x = \nu$. This is the transitional region in which nonoscillatory functions gradually change their behavior and become oscillatory. In accordance with equations (8), (9) and (10) the general solution of Bessel's equation may be expressed as

$$y(x) = \frac{Aw_1(x) + Bw_2(x)}{\sqrt{x/\nu}}, \tag{57}$$

where $w_1(x)$ and $w_2(x)$ constitute a fundamental set of solutions of

$$\frac{d^2w}{dx^2} = \left(-1 + \frac{q^2}{x^2}\right)w, \qquad q^2 = \nu^2 - \frac{1}{4}. \tag{58}$$

The factor $1/\nu$ has been introduced in the denominator of (57) for reasons which will become apparent later. More specifically, we define w_1 and w_2 as follows,

$$\begin{aligned} w_1(q) &= 1, \qquad w_1'(q) = 0; \\ w_2(q) &= 0, \qquad w_2'(q) = 1. \end{aligned} \tag{59}$$

For all solutions of (58)

$$w''(q) = 0; \tag{60}$$

that is, at $x = q$, the curvature of $w(x)$ is zero and in this neighborhood $w(x)$ is represented by $A + B(x - q)$.

To improve on this simple approximation we develop $w(x)$ in a series of powers of $(x - q)$. Thus we obtain

$$\begin{aligned} w_1(x) = 1 &- \frac{(x-q)^3}{3q} + \frac{(x-q)^4}{4q^2} - \frac{(x-q)^5}{5q^3} + \left(\frac{1}{6q^4} + \frac{1}{45q^2}\right)(x-q)^6 \\ &- \left(\frac{1}{7q^5} + \frac{1}{28q^3}\right)(x-q)^7 + \left(\frac{1}{8q^6} + \frac{149}{3360q^4}\right)(x-q)^8 - \cdots, \\ w_2(x) = (x - q) &- \frac{(x-q)^4}{6q} + \frac{3(x-q)^5}{20q^2} - \frac{2(x-q)^6}{15q^3} \\ &+ \left(\frac{5}{42q^4} + \frac{1}{126q^2}\right)(x-q)^7 - \left(\frac{3}{28q^5} + \frac{1}{70q^3}\right)(x-q)^8 + \cdots. \end{aligned} \tag{61}$$

In order to connect this approximate solution with any particular solution for large x we should evaluate the constants A and B in (57) so as to make $y(x)$ assume the required values at two points for which both approximations are valid. Alternatively we can determine A and B so as to match $y(x)$ and

$$y'(x) = \frac{w'(x)}{\sqrt{x/\nu}} - \frac{w(x)}{2x\sqrt{x/\nu}} \tag{62}$$

with the corresponding values of the other approximation for some particular value of x in the region where the approximations overlap.

G. N. Watson's treatise on Bessel functions contains tables of standard Bessel functions and their derivatives at $x = \nu$. Using these values we

can find proper values for A and B to yield either $J_\nu(x)$ or $N_\nu(x)$; thus

$$\begin{aligned} A &= [w_2'(\nu) - \tfrac{1}{2}\nu^{-1}w_2(\nu)]y(\nu) - w_2(\nu)y'(\nu), \\ B &= w_1(\nu)y'(\nu) + [\tfrac{1}{2}\nu^{-1}w_1(\nu) - w_1'(\nu)]y(\nu), \end{aligned} \tag{63}$$

where $y(\nu)$, $y'(\nu)$ represent the values of $J_\nu(\nu)$, $J_\nu'(\nu)$ or of $N_\nu(\nu)$, $N_\nu'(\nu)$. As ν increases we have

$$A \simeq y(\nu) - \tfrac{1}{8}\nu^{-1}y'(\nu), \qquad B = y'(\nu) + \tfrac{1}{2}\nu^{-1}y(\nu). \tag{64}$$

Since $x = 0$ is a singularity of $w(x)$, the power series cannot possibly converge when $|x - q|$ is equal to or exceeds q. In fact the series converge slowly except for small values of $(x - q)/\sqrt[3]{q}$; but they manage to close the gap between the approximations of Liouville's type.

9. *Sinusoidal interpolation and extrapolation*

Consider solutions of the following equation

$$w''(x) = -f(x)w \tag{65}$$

in the interval (a,b). Let

$$\beta^2 = \frac{1}{b-a}\int_a^b f(x)\,dx, \qquad \beta = \sqrt{\frac{1}{b-a}\int_a^b f(x)\,dx}. \tag{66}$$

The first approximation to the solutions is then (see Section 11.13):

$$w(x) = A\sin\beta(x-a) + B\sin\beta(b-x). \tag{67}$$

Letting $x = a, b$ we have

$$w(a) = B\sin\beta(b-a), \qquad w(b) = A\sin\beta(b-a); \tag{68}$$

therefore

$$w(x) = \frac{w(b)\sin\beta(x-a) + w(a)\sin\beta(b-x)}{\sin\beta(b-a)}. \tag{69}$$

This formula may be used for interpolation when $\beta(b-a)$ is not too large, preferably not much larger than $\pi/2$.

If $f(x)$ is negative, then β is imaginary and we may set

$$\gamma^2 = -\frac{1}{b-a}\int_a^b f(x)\,dx, \qquad \gamma = \sqrt{-\frac{1}{b-a}\int_a^b f(x)\,dx}, \tag{70}$$

where γ is real. Then the interpolation formula (69) becomes

$$w(x) = \frac{w(b)\sinh\gamma\,(x-a) + w(a)\sinh\gamma\,(b-x)}{\sinh\gamma\,(b-a)}. \tag{71}$$

For equation (58) we obtain

$$\beta = \sqrt{1 - \frac{q^2}{ab}} = \sqrt{1 - \frac{\nu^2 - 0.25}{ab}},$$
$$\gamma = \sqrt{\frac{q^2}{ab} - 1} = \sqrt{\frac{\nu^2 - 0.25}{ab} - 1}. \tag{72}$$

Taking the values $y(a)$, $y(b)$ of a Bessel function from a table, we can compute $w(a) = \sqrt{a}\,y(a)$, $w(b) = \sqrt{b}\,y(b)$. Substituting these values in (69), we obtain $w(x)$; then $y(x) = w(x)/\sqrt{x}$. The following Table I presents a comparison between the interpolated and exact values for $J_0(x)$ and $J_4(x)$. In the first case $a = 2$, $b = 3$; in the second $a = 3$, $b = 4$.

Table I

x	$y(x)$	$J_0(x)$	*Difference* $y(x) - J_0(x)$
2.0	0.2238908	0.2238908	0
2.1	.1665539	.1666070	−0.0000531
2.2	.1102840	.1103623	− .0000783
2.3	.0554495	.0555398	− .0000903
2.4	.0024102	.0025077	− .0000975
2.5	− .0484869	− .0483838	− .0001031
2.6	− .0969114	− .0968050	− .0001064
2.7	− .1425534	− .1424494	− .0001040
2.8	− .1851262	− .1850360	− .0000902
2.9	− .2243696	− .2243115	− .0000581
3.0	− .2600520	− .2600520	0

x	$y(x)$	$J_4(x)$	*Difference* $y(x) - J_4(x)$
3.0	0.1320342	0.1320342	0
3.1	.1461138	.1456177	0.0004961
3.2	.1602332	.1597218	.0005114
3.3	.1744498	.1742754	.0001744
3.4	.1888187	.1891991	− .0003804
3.5	.2033929	.2044053	− .0010124
3.6	.2182243	.2197990	− .0015747
3.7	.2333637	.2352786	− .0019149
3.8	.2488612	.2507362	− .0018750
3.9	.2647664	.2660587	− .0012923
4.0	.2811291	.2811291	0

If $w(a)$ and $w'(a)$ are known, then

$$\begin{aligned} w(x) &= w(a) \cos \beta(x-a) + \frac{1}{\beta} w'(a) \sin \beta(x-a), \\ w'(x) &= w'(a) \cos \beta(x-a) - \beta w(a) \sin \beta(x-a), \end{aligned} \tag{73}$$

with a corresponding set of hyperbolic formulas when β is imaginary. Since b is at our disposal, we may assume $b = x$ and make β variable. The following Table II compares $J_0(x)$ and the values obtained from (73) when $a = 3$.

Table II

x	3	3.1	3.2	3.3	3.4
$y(x)$	−0.2601	−0.2920	−0.3202	−0.3443	−0.3643
$J_0(x)$	−0.2601	−0.2921	−0.3202	−0.3443	−0.3643

x	3.5	3.6	3.7	3.8	3.9
$y(x)$	−0.3802	−0.3918	−0.3993	−0.4027	−0.4020
$J_0(x)$	−0.3801	−0.3918	−0.3993	−0.4026	−0.4018

Sinusoidal interpolation and extrapolation of solutions of the second order homogeneous differential equations are seen to be very good. The above formulas can be used also if a, b, x, γ are complex.

10. *Recurrence formulas*

Bessel functions and their derivatives of orders $\nu - 1$, ν, $\nu + 1$ are related; thus

$$\begin{aligned}
&\frac{d}{dx}[x^\nu J_\nu(x)] = x^\nu J_{\nu-1}(x), && \frac{d}{dx}[x^\nu N_\nu(x)] = x^\nu N_{\nu-1}(x); \\
&\frac{d}{dx}[x^{-\nu} J_\nu(x)] = -x^{-\nu} J_{\nu+1}(x), && \frac{d}{dx}[x^{-\nu} N_\nu(x)] = -x^{-\nu} N_{\nu+1}(x); \\
&xJ_\nu' + \nu J_\nu = xJ_{\nu-1}, && xN_\nu' + \nu N_\nu = xN_{\nu-1}; \\
&xJ_\nu' - \nu J_\nu = -xJ_{\nu+1}, && xN_\nu' - \nu N_\nu = -xN_{\nu+1}; \\
&J_{\nu-1} + J_{\nu+1} = \frac{2\nu}{x} J_\nu, && N_{\nu-1} + N_{\nu+1} = \frac{2\nu}{x} N_\nu; \\
&J_{\nu-1} - J_{\nu+1} = 2J_\nu', && N_{\nu-1} - N_{\nu+1} = 2N_\nu'.
\end{aligned} \tag{74}$$

The first and third formulas are proved by substituting the power series for J_ν and differentiating; thus

$$\begin{aligned}\frac{d}{dx}[x^\nu J_\nu] &= \frac{d}{dx}\sum_{m=0}^{\infty}\frac{(-)^m x^{2\nu+2m}}{2^{\nu+2m}m!(\nu+m)!}\\ &= \sum_{m=0}^{\infty}\frac{(-)^m x^{2\nu-1+2m}}{2^{\nu-1+2m}m!(\nu-1+m)!} = x^\nu J_{\nu-1},\end{aligned} \tag{75}$$

and similarly for the third formula.

Since the power series for J_ν is valid for all values of ν so are the formulas. Replacing ν by $-\nu$ and using the definition for N_ν, the second and fourth formulas are found from the proper combination of J_ν and $J_{-\nu}$. The next two lines in (74) follow from the first two when the indicated differentiation has been performed. Adding and subtracting these two lines, we obtain the last two.

A similar set of formulas may be obtained for the modified Bessel functions:

$$\begin{aligned}
&\frac{d}{dx}[x^\nu I_\nu(x)] = x^\nu I_{\nu-1}(x), && \frac{d}{dx}[x^\nu K_\nu(x)] = -x^\nu K_{\nu-1}(x);\\
&\frac{d}{dx}[x^{-\nu} I_\nu(x)] = x^{-\nu} I_{\nu+1}(x), && \frac{d}{dx}[x^{-\nu} K_\nu(x)] = -x^{-\nu} K_{\nu+1}(x);\\
&xI_\nu' + \nu I_\nu = xI_{\nu-1}, && xK_\nu' + \nu K_\nu = -xK_{\nu-1};\\
&xI_\nu' - \nu I_\nu = xI_{\nu+1}, && xK_\nu' - \nu K_\nu = -xK_{\nu+1};\\
&I_{\nu-1} - I_{\nu+1} = \frac{2\nu}{x} I_\nu, && K_{\nu-1} - K_{\nu+1} = -\frac{2\nu}{x} K_\nu;\\
&I_{\nu-1} + I_{\nu+1} = 2I_\nu', && K_{\nu-1} + K_{\nu+1} = -2K_\nu'.
\end{aligned} \tag{76}$$

Thus Bessel functions of higher orders may be computed from those of lower orders.

11. *A connection between Bessel functions and their derivatives*

Bessel's equation may be expressed in the form (11–94) with $P(x) = x$ and $Q(x) = (\nu^2/x) - x$. Applying (11–95), we have

$$y_2(x)y_1'(x) - y_1(x)y_2'(x) = C/x \tag{77}$$

for any pair of solutions. The constant C may be calculated at some convenient point such as $x = 0$ or $x = \infty$. Thus for the pair J_ν, N_ν we have

$$J_\nu N_\nu' - J_\nu' N_\nu = 2/\pi x. \tag{78}$$

If x is a zero of J_ν, then $J'_\nu = -2/\pi x N_\nu$. Similar identities hold for the zeros of other functions in (78).

For the modified Bessel functions we have

$$K_\nu I'_\nu - K'_\nu I_\nu = 1/x. \tag{79}$$

For any pair of solutions of the reduced equation (58) we have

$$w_2 w'_1 - w_1 w'_2 = C. \tag{80}$$

If w_1 and w_2 behave as $\sin x$ and $\cos x$ at infinity, then $C = 1$.

12. *Integrals of Bessel functions and their products*

Some integrals containing Bessel functions can be obtained directly from (74); thus

$$\begin{aligned} &\int x^\nu J_{\nu-1}(x)\,dx = x^\nu J_\nu(x), &&\int x^\nu N_{\nu-1}(x) = x^\nu N_\nu(x); \\ &\int x^{-\nu} J_{\nu+1}(x)\,dx = -x^{-\nu} J_\nu(x), &&\int x^{-\nu} N_{\nu+1}(x) = -x^{-\nu} N_\nu(x). \end{aligned} \tag{81}$$

A variety of integrals involving products of Bessel functions may be evaluated as follows. Starting with the equations

$$\frac{d}{dx}\left(x\frac{dy_\nu}{dx}\right) = \left(\frac{\nu^2}{x} - x\right)y_\nu, \quad \frac{d}{dx}\left(x\frac{dy_\mu}{dx}\right) = \left(\frac{\mu^2}{x} - x\right)y_\mu, \tag{82}$$

we multiply the first by y_μ, the second by y_ν and subtract

$$\frac{d}{dx}\left(xy_\mu\frac{dy_\nu}{dx} - xy_\nu\frac{dy_\mu}{dx}\right) = \frac{\nu^2 - \mu^2}{x}y_\mu y_\nu. \tag{83}$$

Integrating

$$(\nu^2 - \mu^2)\int x^{-1} y_\mu y_\nu\,dx = x\left(y_\mu\frac{dy_\nu}{dx} - y_\nu\frac{dy_\mu}{dx}\right). \tag{84}$$

Let us now make $\mu = \nu$ and replace x by βx in the first equation in (82) and by kx in the second:

$$\begin{aligned} \frac{d}{dx}\left[x\frac{dy_\nu(\beta x)}{dx}\right] &= \left(\frac{\nu^2}{x} - \beta^2 x\right)y_\nu(\beta x), \\ \frac{d}{dx}\left[x\frac{d\bar{y}_\nu(kx)}{dx}\right] &= \left(\frac{\nu^2}{x} - k^2 x\right)\bar{y}_\nu(kx). \end{aligned} \tag{85}$$

Multiplying the first equation by $\bar{y}_\nu(kx)$, the second by $y_\nu(\beta x)$, then sub-

tracting and integrating

$$(k^2 - \beta^2)\int_a^x xy_\nu(\beta x)\bar{y}_\nu(kx)\,dx$$

$$= x\left[\bar{y}_\nu(kx)\frac{dy_\nu(\beta x)}{dx} - y_\nu(\beta x)\frac{d\bar{y}_\nu(kx)}{dx}\right]\Bigg|_a^x. \quad (86)$$

The bar is used to indicate that as $k \to \beta$, $\bar{y}_\nu(kx)$ need not approach $y_\nu(\beta x)$.

If we divide (86) by $k^2 - \beta^2$ and let k approach β, we obtain

$$\int_a^x xy_\nu(\beta x)\bar{y}_\nu(\beta x)\,dx = \frac{x}{2\beta}\frac{\partial}{\partial k}\left[\bar{y}_\nu(kx)\frac{dy_\nu(\beta x)}{dx} - y_\nu(\beta x)\frac{d\bar{y}_\nu(kx)}{dx}\right]_{k=\beta}$$

$$= \frac{1}{2}x^2\left[y_\nu'\bar{y}_\nu' + \left(1 - \frac{\nu^2}{\beta^2x^2}\right)y_\nu\bar{y}_\nu\right]\Bigg|_a^x \quad (87)$$

where *the primes denote differentiation with respect to* βx. *In particular* both y_ν and $\bar{y}_\nu$ could be equal to the same Bessel function, $J_\nu(\beta x)$ for example.

13. *Orthogonal expansions*

Certain sets of Bessel functions are orthogonal and therefore suitable for expansions of arbitrary functions. Consider the following Bessel equation

$$\frac{d}{dx}\left(x\frac{dy_\nu}{dx}\right) = \left(\frac{\nu^2}{x} - \chi^2 x\right)y_\nu \quad (88)$$

and its general solution

$$y_\nu(\chi x) = AJ_\nu(\chi x) + BN_\nu(\chi x). \quad (89)$$

If k and β are two distinct values of χ for which the right side in (86) vanishes at $x = a$ and $x = b$, then the corresponding y-functions are orthogonal with the weight function equal to x; thus

$$\int_a^b xy_\nu(\chi_1 x)y_\nu(\chi_2 x)\,dx = 0, \qquad \chi_1 \neq \chi_2. \quad (90)$$

The right side in (86) is certainly zero if $y_\nu(\chi x)$ or its derivative vanishes at $x = a, b$; that is, if the χ's are roots of any of the following equations

$$\frac{J_\nu(\chi a)}{N_\nu(\chi a)} = \frac{J_\nu(\chi b)}{N_\nu(\chi b)}, \qquad \frac{J_\nu'(\chi a)}{N_\nu'(\chi a)} = \frac{J_\nu'(\chi b)}{N_\nu'(\chi b)},$$

$$\frac{J_\nu(\chi a)}{N_\nu(\chi a)} = \frac{J_\nu'(\chi b)}{N_\nu'(\chi b)}, \qquad \frac{J_\nu'(\chi a)}{N_\nu'(\chi a)} = \frac{J_\nu(\chi b)}{N_\nu(\chi b)}. \quad (91)$$

More generally we can add and subtract $P\bar{y}_\nu(kx)y_\nu(\beta x)$ on the right in (86) to obtain

$$\bar{y}_\nu(kx)\left[x\frac{dy_\nu(\beta x)}{dx} + Py_\nu(\beta x)\right] - y_\nu(\beta x)\left[x\frac{dy_\nu(kx)}{dx} + Py_\nu(kx)\right]. \quad (92)$$

This expression will vanish if k and β are any two distinct roots of

$$a\frac{dy_\nu(\chi a)}{da} + P_1y_\nu(\chi a) = 0,$$

$$b\frac{dy_\nu(\chi b)}{db} + P_2y_\nu(\chi b) = 0, \quad (93)$$

where P_1 and P_2 are any two given constants. We may let P_1 and P_2 assume such values as 0, 1, ∞ and thus obtain the special cases (91).

Therefore, if

$$f(x) = \sum_m a_m y_\nu(\chi_m x), \quad (94)$$

then

$$a_m = \frac{\int_a^b xf(x)y_\nu(\chi_m x)\,dx}{\int_a^b x[y_\nu(\chi_m x)]^2\,dx}. \quad (95)$$

The denominator is given by (87) with $\bar{y}_\nu = y_\nu$.

14. *Zeros of Bessel functions*

Approximate zeros of $J_\nu(x)$ and $N_\nu(x)$ for $\nu > 1/2$ may be obtained from the approximate formula (45) for these functions. These zeros are solutions of the following equation

$$\sqrt{(x/q)^2 - 1} - \tan^{-1}\sqrt{(x/q)^2 - 1} = \frac{\vartheta + \frac{1}{4}\pi - \frac{1}{2}(q - \nu)\pi}{q}. \quad (96)$$

For the successive zeros of $J_\nu(x)$, $\vartheta = \pi/2, 3\pi/2, 5\pi/2, \cdots$; for the successive zeros of $N_\nu(x)$, $\vartheta = 0, \pi, 2\pi, \cdots$. The following tables compare the approximate and exact values for $\nu = 1, 2$.

Zeros of $J_1(x)$			
exact	3.8317	7.0156	10.1735
appr.	3.8286	7.0151	10.1733
difference	0.0031	0.0005	0.0002

Zeros of $N_1(x)$			
exact	2.1971	5.4297	8.5960
appr.	2.1820	5.4286	8.5957
difference	0.0152	0.0011	0.0003

	Zeros of $J_2(x)$				Zeros of $N_2(x)$		
exact	5.1356	8.4172	11.6198	exact	3.3842	6.7938	10.0235
appr.	5.1276	8.4156	11.6197	appr.	3.3497	6.7905	10.0225
difference	0.0080	0.0016	0.0001	difference	0.0345	0.0033	0.0010

15. *Bessel functions in the complex plane*

In the complex plane Bessel functions are generally multiple-valued functions. From the power series for J_ν, $J_{-\nu}$, I_ν, $I_{-\nu}$, we obtain

$$\begin{aligned} J_\nu(ze^{im\pi}) &= e^{im\nu\pi}J_\nu(z), \qquad J_{-\nu}(ze^{im\pi}) = e^{-im\nu\pi}J_{-\nu}(z), \\ I_\nu(ze^{im\pi}) &= e^{im\nu\pi}I_\nu(z), \qquad I_{-\nu}(ze^{im\pi}) = e^{-im\nu\pi}I_{-\nu}(z), \end{aligned} \tag{97}$$

where m is an integer. Then from the definitions of the N, H, and K functions,

$$\begin{aligned} N_\nu(ze^{im\pi}) &= e^{-im\nu\pi}N_\nu(z) + 2i\sin m\nu\pi\cot\nu\pi J_\nu(z), \\ H_\nu^{(1)}(ze^{im\pi}) &= \frac{\sin(1-m)\nu\pi}{\sin\nu\pi}H_\nu^{(1)}(z) - e^{-i\nu\pi}\frac{\sin m\nu\pi}{\sin\nu\pi}H_\nu^{(2)}(z), \\ H_\nu^{(2)}(ze^{im\pi}) &= \frac{\sin(1+m)\nu\pi}{\sin\nu\pi}H_\nu^{(2)}(z) + e^{i\nu\pi}\frac{\sin m\nu\pi}{\sin\nu\pi}H_\nu^{(1)}(z), \\ K_\nu(ze^{im\pi}) &= e^{-im\nu\pi}K_\nu(z) - i\pi\frac{\sin m\nu\pi}{\sin\nu\pi}I_\nu(z). \end{aligned} \tag{98}$$

Approximations of Liouville's type are also multiple-valued functions; but their branches do not correspond to the branches of Bessel functions. The same approximation will represent one Bessel function in one part of the complex plane and another function in another part.

16. *Miscellaneous formulas*

The following are some of the more frequently useful formulas in addition to those given in the preceding sections:

$$\begin{aligned} J_0'(x) &= -J_1(x), \qquad N_0'(x) = -N_1(x); \\ I_0'(x) &= I_1(x), \qquad K_0'(x) = -K_1(x). \end{aligned} \tag{99}$$

$$\int_0^x xJ_\nu^2(kx)\,dx = \tfrac{1}{2}x^2[J_\nu^2(kx) - J_{\nu-1}(kx)J_{\nu+1}(kx)], \quad \text{see (87)}. \tag{100}$$

$$\int_0^x xy_\nu^2(kx)\,dx = \tfrac{1}{2}x^2[y_\nu^2(kx) - y_{\nu-1}(kx)y_{\nu+1}(kx)], \quad \text{see (87)}. \tag{101}$$

If $J_\nu(k) = 0$, then
$$J'_\nu(k) = J_{\nu-1}(k) = -J_{\nu+1}(k), \qquad J''_\nu(k) = -(1/k)J'_\nu(k). \tag{102}$$

If $N_\nu(k) = 0$, then
$$N'_\nu(k) = N_{\nu-1}(k) = -N_{\nu+1}(k), \qquad N''_\nu(k) = -(1/k)N'_\nu(k). \tag{103}$$

If $J'_\nu(k) = 0$, then
$$J_{\nu+1}(k) = J_{\nu-1}(k) = (\nu/k)J_\nu(k), \qquad J''_\nu(k) = (\nu^2/k^2 - 1)J_\nu(k). \tag{104}$$

If $N'_\nu(k) = 0$, then
$$N_{\nu+1}(k) = N_{\nu-1}(k) = (\nu/k)N_\nu(k), \qquad N''_\nu(k) = (\nu^2/k^2 - 1)N_\nu(k). \tag{105}$$

If $I_\nu(k) = 0$, then
$$I'_\nu(k) = I_{\nu-1}(k) = I_{\nu+1}(k), \qquad I''_\nu(k) = -(1/k)I'_\nu(k). \tag{106}$$

If $K_\nu(k) = 0$, then
$$K'_\nu(k) = -K_{\nu-1}(k) = -K_{\nu+1}(k), \qquad K''_\nu(k) = -(1/k)K'_\nu(k). \tag{107}$$

If $I'_\nu(k) = 0$, then
$$I_{\nu-1}(k) = -I_{\nu+1}(k) = (\nu/k)I_\nu(k), \qquad I''_\nu(k) = (\nu^2/k^2 + 1)I_\nu(k). \tag{108}$$

If $K'_\nu(k) = 0$, then
$$K_{\nu+1}(k) = -K_{\nu-1}(k) = (\nu/k)K_\nu(k), \quad K''_\nu(k) = (\nu^2/k^2 + 1)K_\nu(k). \tag{109}$$

$$x\hat{J}'_\nu = (\nu + 1)\hat{J}_\nu - x\hat{J}_{\nu+1}, \qquad x\hat{N}'_\nu = (\nu + 1)\hat{N}_\nu - x\hat{N}_{\nu+1}. \tag{110}$$

$$x\hat{J}'_\nu = -\nu\hat{J}_\nu + x\hat{J}_{\nu-1}, \qquad x\hat{N}'_\nu = -\nu\hat{N}_\nu + x\hat{N}_{\nu-1}. \tag{111}$$

$$x\hat{I}'_\nu = (\nu + 1)\hat{I}_\nu + x\hat{I}_{\nu+1}, \qquad x\hat{K}'_\nu = (\nu + 1)\hat{K}_\nu - x\hat{K}_{\nu+1}. \tag{112}$$

$$x\hat{I}'_\nu = -\nu\hat{I}_\nu + x\hat{I}_{\nu-1}, \qquad x\hat{K}'_\nu = -\nu\hat{K}_\nu - x\hat{K}_{\nu-1}. \tag{113}$$

If $\hat{J}_\nu(k) = 0$, then
$$\hat{J}'_\nu(k) = \hat{J}_{\nu-1}(k) = -\hat{J}_{\nu+1}(k), \qquad \hat{J}''_\nu(k) = 0. \tag{114}$$

If $\hat{N}_\nu(k) = 0$, then
$$\hat{N}'_\nu(k) = \hat{N}_{\nu-1}(k) = -\hat{N}_{\nu+1}(k), \qquad \hat{N}''_\nu(k) = 0. \tag{115}$$

If $\hat{J}'_\nu(k) = 0$, then
$$\begin{aligned} \hat{J}_{\nu+1} &= (\nu + 1)\hat{J}_\nu/k, \qquad \hat{J}_{\nu-1} = \nu\hat{J}_\nu/k, \\ \hat{J}''_\nu &= [\nu(\nu + 1)/k^2 - 1]\hat{J}_\nu. \end{aligned} \tag{116}$$

If $\hat{N}'_\nu(k) = 0$, then

$$\begin{aligned} \hat{N}_{\nu+1} &= (\nu + 1)\hat{N}_\nu/k, \qquad \hat{N}_{\nu-1} = \nu\hat{N}_\nu/k, \\ \hat{N}''_\nu &= [\nu(\nu + 1)/k^2 - 1]\hat{N}_\nu. \end{aligned} \tag{117}$$

If $\hat{I}_\nu(k) = 0$, then $\hat{I}'_\nu = \hat{I}_{\nu-1} = \hat{I}_{\nu+1}, \qquad \hat{I}''_\nu = 0.$ (118)

If $\hat{K}_\nu(k) = 0$, then $\hat{K}'_\nu = -\hat{K}_{\nu-1} = -\hat{K}_{\nu+1}, \qquad \hat{K}''_\nu = 0.$ (119)

If $\hat{I}'_\nu(k) = 0$, then

$$\begin{aligned} \hat{I}_{\nu+1} &= -(\nu + 1)\hat{I}_\nu/k, \qquad \hat{I}_{\nu-1} = \nu\hat{I}_\nu/k, \\ \hat{I}''_\nu &= [\nu(\nu + 1)/k^2 + 1]\hat{I}_\nu. \end{aligned} \tag{120}$$

If $\hat{K}'_\nu(k) = 0$, then

$$\begin{aligned} \hat{K}_{\nu+1} &= (\nu + 1)\hat{K}_\nu/k, \qquad \hat{K}_{\nu-1} = -\nu\hat{K}_\nu/k, \\ \hat{K}''_\nu &= [\nu(\nu + 1)/k^2 + 1]\hat{K}_\nu. \end{aligned} \tag{121}$$

$$\begin{aligned} K_\nu(ix) &= \tfrac{1}{2}\pi e^{-i(\nu+1)\pi/2}[J_\nu(x) - iN_\nu(x)], \\ K'_\nu(ix) &= \tfrac{1}{2}\pi e^{-i(\nu+1)\pi/2}[-iJ'_\nu(x) - N'_\nu(x)]. \end{aligned} \tag{122}$$

$$\hat{I}_\nu(ix) = e^{i(\nu+1)\pi/2}\hat{J}_\nu(x), \qquad \hat{I}'_\nu(ix) = e^{i\nu\pi/2}\hat{J}'_\nu(x). \tag{123}$$

$$\begin{aligned} \hat{K}_\nu(ix) &= e^{-i(\nu+1)\pi/2}[\hat{J}_\nu(x) - i\hat{N}_\nu(x)], \\ \hat{K}'_\nu(ix) &= e^{-i(\nu+1)\pi/2}[-i\hat{J}'_\nu(x) - \hat{N}'_\nu(x)]. \end{aligned} \tag{124}$$

$$\hat{J}_\nu(x)\hat{N}'_\nu(x) - \hat{J}'_\nu(x)\hat{N}_\nu(x) = \hat{J}_{\nu+1}\hat{N}_\nu - \hat{J}_\nu\hat{N}_{\nu+1} = 1. \tag{125}$$

$$\frac{K'_\nu(ix)}{K_\nu(ix)} = -\frac{2}{\pi x[J_\nu^2(x) + N_\nu^2(x)]} - i\,\frac{J_\nu J'_\nu + N_\nu N'_\nu}{J_\nu^2 + N_\nu^2}. \tag{126}$$

$$\frac{K_\nu(ix)}{K'_\nu(ix)} = -\frac{2}{\pi x[(J'_\nu)^2 + (N'_\nu)^2]} + i\,\frac{J_\nu J'_\nu + N_\nu N'_\nu}{(J'_\nu)^2 + (N'_\nu)^2}. \tag{127}$$

$$\frac{\hat{K}'_\nu(ix)}{\hat{K}_\nu(ix)} = -\frac{1}{\hat{J}_\nu^2(x) + \hat{N}_\nu^2(x)} - i\,\frac{\hat{J}_\nu\hat{J}'_\nu + \hat{N}_\nu\hat{N}'_\nu}{\hat{J}_\nu^2 + \hat{N}_\nu^2}. \tag{128}$$

$$\frac{\hat{K}_\nu(ix)}{\hat{K}'_\nu(ix)} = -\frac{1}{(\hat{J}'_\nu)^2 + (\hat{N}'_\nu)^2} + i\,\frac{\hat{J}_\nu\hat{J}'_\nu + \hat{N}_\nu\hat{N}'_\nu}{(\hat{J}'_\nu)^2 + (\hat{N}'_\nu)^2}. \tag{129}$$

$$J_0^2 + N_0^2 \sim \frac{2}{\pi x}\left[1 + \sum_{m=1}^{\infty} \frac{(-)^m[1\cdot 3\cdot 5\cdots(2m-1)]^4}{(2m)!(2x)^{2m}}\right]. \tag{130}$$

$$J_\nu^2 + N_\nu^2 \sim \frac{2}{\pi x}\left[1 + \sum_{m=1}^{\infty} [1\cdot 3\cdot 5\cdots(2m-1)]\frac{(\nu,m)}{2^m x^{2m}}\right]. \tag{131}$$

$$\begin{aligned}(\nu,m) &= \frac{(\nu+m-\frac{1}{2})!}{m!(\nu-m-\frac{1}{2})!}\\ &= \frac{(4\nu^2-1^2)(4\nu^2-3^2)\cdots[4\nu^2-(2m-1)^2]}{2^{2m}m!}.\end{aligned} \tag{132}$$

$$\hat{J}_\nu^2 + \hat{N}_\nu^2 \sim 1 + \sum_{m=1}^{\infty}\left[\frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdots\left(m-\frac{1}{2}\right)\right]\frac{(\nu+m)!}{m!(\nu-m)!x^{2m}} \tag{133}$$

$$\hat{J}_n^2 + \hat{N}_n^2 = 1 + \sum_{m=1}^{n}\left[\frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdots\left(m-\frac{1}{2}\right)\right]\frac{(n+m)!}{m!(n-m)!x^{2m}}. \tag{134}$$

$$\hat{J}_0^2 + \hat{N}_0^2 = 1 \tag{135}$$

$$\int_0^\infty J_n(x)\,dx = 1, \qquad \int_0^\infty \frac{J_n(kx)}{x}\,dx = \frac{1}{n}, \qquad n = 1,2,3,\cdots. \tag{136}$$

$$\int_0^x \frac{J_1(x)}{x}\,dx \sim 1 - \sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2n-3)J_{n-1}(x)}{x^n}. \tag{137}$$

$$\begin{aligned}\int_x^\infty \frac{J_n(x)}{x^m}\,dx \sim &-\frac{J_{n+1}(x)}{x^m} - \frac{(m+n+1)J_{n+2}(x)}{x^{m+1}}\\ &- \frac{(m+n+1)(m+n+3)J_{n+3}(x)}{x^{m+2}} - \cdots.\end{aligned} \tag{138}$$

$$\begin{aligned}\exp(\pm i\beta\rho\sin\varphi) = J_0(\beta\rho) &+ 2\sum_{n=1}^{\infty} J_{2n}(\beta\rho)\cos 2n\varphi\\ &\pm 2i\sum_{n=0}^{\infty} J_{2n+1}(\beta\rho)\sin(2n+1)\varphi.\end{aligned} \tag{139}$$

$$\begin{aligned}\exp(\pm i\beta\rho\cos\varphi) = J_0(\beta\rho) &+ 2\sum_{n=1}^{\infty}(-)^n J_{2n}(\beta\rho)\cos 2n\varphi\\ &\pm 2i\sum_{n=0}^{\infty}(-)^n J_{2n+1}(\beta\rho)\cos(2n+1)\varphi.\end{aligned} \tag{140}$$

$$\exp(\pm\sigma\rho\cos\varphi) = I_0(\sigma\rho) + 2\sum_{n=1}^{\infty} I_n(\pm\sigma\rho)\cos n\varphi. \tag{141}$$

$$\exp(\pm\sigma\rho\sin\varphi) = I_0(\sigma\rho) + 2\sum_{n=1}^{\infty}(-)^n I_{2n}(\sigma\rho)\cos 2n\varphi$$
$$\pm 2\sum_{n=0}^{\infty}(-)^n I_{2n+1}(\sigma\rho)\sin(2n+1)\varphi. \tag{142}$$

$$K_0(\sigma R) = K_0(\sigma\rho_0)I_0(\sigma\rho) + 2\sum_{n=1}^{\infty} K_n(\sigma\rho_0)I_n(\sigma\rho)\cos n(\varphi-\varphi_0),$$
$$\rho < \rho_0,$$
$$= I_0(\sigma\rho_0)K_0(\sigma\rho) + 2\sum_{n=1}^{\infty} I_n(\sigma\rho_0)K_n(\sigma\rho)\cos n(\varphi-\varphi_0), \tag{143}$$
$$\rho > \rho_0,$$
$$R = \sqrt{\rho^2 - 2\rho\rho_0\cos(\varphi-\varphi_0) + \rho_0^2}, \qquad \mathrm{re}\,(\sigma) \geq 0.$$

$$\tfrac{1}{2}[f(\rho+0) + f(\rho-0)] = \sum_m a_m J_n(k_m\rho/a), \quad \text{where}$$
$$Ak_m J_n'(k_m) + BJ_n(k_m) = 0,$$
$$a_m = \frac{2k_m^2\int_0^a \rho f(\rho)J_n(k_m\rho/a)\,d\rho}{a^2\{[k_m J_n'(k_m)]^2 + (k_m^2 - n^2)[J_n(k_m)]^2\}}$$
$$= \frac{2A^2k_m^2\int_0^a \rho f(\rho)J_n(k_m\rho/a)\,d\rho}{a^2[B^2 + A^2(k_m^2 - n^2)]J_n^2(k_m)} \tag{144}$$
$$= \frac{2B^2\int_0^a \rho f(\rho)J_n(k_m\rho/a)\,d\rho}{a^2[B^2 + A^2(k_m^2 - n^2)][J_n'(k_m)]^2}.$$

$$\int_0^\infty d\chi\int_a^b \chi\hat{\rho}f(\rho)J_\nu(\chi\hat{\rho})J_\nu(\chi\rho)\,d\hat{\rho}$$
$$= \tfrac{1}{2}[f(\rho+0)+f(\rho-0)], \qquad a < \rho < b,$$
$$= \tfrac{1}{2}f(\rho+0), \qquad \rho = a, \tag{145}$$
$$= \tfrac{1}{2}f(\rho-0), \qquad \rho = b,$$
$$= 0, \qquad 0 < \rho < a \quad \text{or} \quad \rho > b, \qquad \mathrm{re}\,\nu > -1.$$

If $\dfrac{d^2w}{dx^2} = -\alpha^2 w + \dfrac{\nu(\nu+1)}{x^2}\,w, \qquad \dfrac{d^2\bar{w}}{dx^2} = -\beta^2\bar{w} + \dfrac{\nu(\nu+1)}{x^2}\,\bar{w},$ (146)

then $$(\alpha^2 - \beta^2)\int^x w\bar{w}\,dx = w\frac{d\bar{w}}{dx} - \bar{w}\frac{dw}{dx}.$$

If $\alpha = \beta$, then

$$\int^x w_1 w_2\,dx = \tfrac{1}{2}x\left\{w_1' w_2' + \left[1 - \frac{\nu(\nu+1)}{\beta^2 x^2}\right] w_1 w_2 - \frac{1}{2\beta x}\,(w_1' w_2 + w_1 w_2')\right\}, \tag{147}$$

where the primes denote differentiation with respect to βx.

CHAPTER XXI

Legendre Functions

1. *Legendre's equation*

Legendre's equation,

$$\frac{d^2\Theta}{d\theta^2} + \cot\theta \frac{d\Theta}{d\theta} + \nu(\nu + 1)\Theta = 0, \tag{1}$$

plays an important role in electrostatics and in the theory of spherical waves. Its solutions are called *Legendre functions.* Another name for the functions is *zonal harmonics.* The parameter ν may be either real or complex. In the theory of spherical waves in free space ν is an integer; in the theory of nonabsorbing conical sound horns and electric horns ν is either an integer or a fraction; in the theory of absorbing horns it is a complex number.

Another common form of Legendre's equation is obtained if we change the independent variable by the substitution

$$x = \cos\theta; \tag{2}$$

thus

$$(1 - x^2)\frac{d^2\Theta}{dx^2} - 2x\frac{d\Theta}{dx} + \nu(\nu + 1)\Theta = 0. \tag{3}$$

In many applications θ is an angle (in spherical coordinates the polar angle or colatitude) varying from zero to π; then x lies in the interval $(-1, 1)$. In other applications x is greater than unity and θ is imaginary.

Before considering the standard forms of Legendre functions let us examine the behavior of solutions of (1) for real values of θ. If θ is near $\pi/2$, $\cot\theta$ is small and Θ is approximately a sinusoidal function. Let us remove the term containing the first derivative (see Section 11.7) by substituting

$$\Theta = \hat{\Theta}/\sqrt{\sin\theta}. \tag{4}$$

Thus we obtain

$$\begin{aligned} \frac{d^2\hat{\Theta}}{d\theta^2} &= -\left[(\nu + \tfrac{1}{2})^2 + \tfrac{1}{4}\csc^2\theta\right]\hat{\Theta} \\ &= -\left[(\nu + \tfrac{1}{2})^2 + \tfrac{1}{4} + \tfrac{1}{4}\cot^2\theta\right]\hat{\Theta}. \end{aligned} \tag{5}$$

In a θ-interval in which the last term in the brackets is small in comparison with the first two, we have approximately

$$\hat{\Theta} = A \cos k\theta + B \sin k\theta, \qquad k = \sqrt{(\nu + \tfrac{1}{2})^2 + \tfrac{1}{4}}\,;$$
$$\Theta = \frac{A \cos k\theta + B \sin k\theta}{\sqrt{\sin \theta}}. \tag{6}$$

The interval in which this approximation is valid increases as ν increases.

If θ is small, $\cot \theta \simeq 1/\theta$ and the Legendre equation is approximated by the Bessel equation of order zero (the standard form is obtained if we let $\theta = u/\sqrt{\nu(\nu + 1)}$); hence for small θ we have

$$\Theta = A' J_0(\beta\theta) + B' N_0(\beta\theta), \qquad \beta = \sqrt{\nu(\nu + 1)}\,. \tag{7}$$

A similar approximation is obtained for θ nearly equal to π if we replace θ in (7) by $\pi - \theta$. In fact, *if* $\Theta(\theta)$ *is a solution of* (1), *then* $\Theta(\pi - \theta)$ *is also a solution.*

Thus there exists one type of solution which reduces to a constant at $\theta = 0$ and another type which varies as $\log \theta$ for small θ. Likewise some solutions reduce to a constant at $\theta = \pi$ and others vary as $\log (\pi - \theta)$ for small $(\pi - \theta)$. Standard forms of Legendre functions are usually written as functions of $x = \cos \theta$; the one which reduces to unity at $\theta = 0$, $x = 1$ is denoted by $P_\nu (\cos \theta)$. The solution $P_\nu[\cos (\pi - \theta)] = P_\nu(-\cos \theta)$ reduces to unity when $\theta = \pi$; for all nonintegral values of ν this solution is different from $P_\nu (\cos \theta)$ and therefore

$$\Theta = A'' P_\nu (\cos \theta) + B'' P_\nu (-\cos \theta) \tag{8}$$

is the general solution of Legendre's equation.

From the definition of P_ν and the behavior of Θ for small θ we have

$$P_\nu (\cos \theta) \simeq J_0(\beta\theta) = J_0(\sqrt{\nu(\nu + 1)}\,\theta). \tag{9}$$

If ν is large, the argument of the Bessel function may be large even when θ itself is small. Replacing the Bessel function by its asymptotic approximation, we obtain

$$P_\nu (\cos \theta) \sim \frac{\cos\left(\beta\theta - \dfrac{\pi}{4}\right)}{\sqrt{\tfrac{1}{2}\pi\beta\theta}}. \tag{10}$$

If ν is large, the difference between k in (6) and β in (7) is small,

$$k - \beta \simeq \frac{1}{2(2\nu + 1)}, \qquad k\theta - \beta\theta \simeq \frac{\theta}{2(2\nu + 1)}. \tag{11}$$

For small θ, $\sin\theta \simeq \theta$; thus in view of equation (6) we can extend (10) to larger values of θ

$$P_\nu(\cos\theta) \sim \frac{\cos\left(k\theta - \frac{\pi}{4}\right)}{\sqrt{\frac{1}{2}\pi\beta\sin\theta}}. \tag{12}$$

As ν increases, β and k approach $\nu + \frac{1}{2}$.

2. *Power series for $P_\nu(x)$*

Since $P_\nu(x)$, where $x = \cos\theta$, reduces to unity at $x = 1$, we shall seek an expression for it in powers of $(1 - x)$. For this purpose we express the coefficients in (3) as follows: $1 - x^2 = 2(1 - x) - (1 - x)^2$, $-2x = 2(1 - x) - 2$; then we substitute for Θ a series of the form $1 + a_1(1 - x) + a_2(1 - x)^2 + \cdots$. Equating the coefficients of various powers of $(1 - x)$ to zero, we obtain

$$\begin{aligned} P_\nu(x) = 1 - \nu(\nu+1)\frac{1-x}{2} &+ \frac{\nu(\nu-1)(\nu+1)(\nu+2)}{(2!)^2}\left(\frac{1-x}{2}\right)^2 \\ &- \frac{\nu(\nu-1)(\nu-2)(\nu+1)(\nu+2)(\nu+3)}{(3!)^2}\left(\frac{1-x}{2}\right)^3 + \cdots. \end{aligned} \tag{13}$$

In terms of θ, $(1 - x)/2 = \sin^2(\theta/2)$.

This series is a special case of the *hypergeometric function:*

$$\begin{aligned} F(\alpha,\beta;\gamma;z) &= 1 + \frac{\alpha\cdot\beta}{1\cdot\gamma}z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}z^2 + \cdots \\ &= 1 + \frac{(\gamma-1)!}{(\alpha-1)!(\beta-1)!}\sum_{m=0}^{\infty}\frac{(\alpha+m)!(\beta+m)!}{m!(\gamma+m)!}z^{m+1}. \end{aligned} \tag{14}$$

This function becomes $P_\nu(x)$ if

$$\alpha = -\nu, \qquad \beta = \nu + 1, \qquad \gamma = 1, \qquad z = \frac{1-x}{2}. \tag{15}$$

The hypergeometric series is convergent if $|z| < 1$. Thus (13) is valid in the interval $-1 < x \leq 1$, $0 \leq \theta < \pi$. It can be shown that as x approaches negative unity and θ approaches π, the series behaves ultimately as $\log\frac{1}{2}(1 + x) = 2\log\cos(\theta/2)$.

The series for $P_\nu(x)$ may be written as follows:

$$P_\nu(x) = \sum_{m=0}^{n} \frac{(-)^m(\nu+m)!}{m!\,m!(\nu-m)!}\left(\frac{1-x}{2}\right)^m$$
$$-\frac{\sin \nu\pi}{\pi}\sum_{m=n+1}^{\infty}\frac{(m-1-\nu)!(m+\nu)!}{m!\,m!}\left(\frac{1-x}{2}\right)^m, \tag{16}$$

where $\nu = n + \delta$, n is an integer, and δ is a proper fraction. The factor $\sin \nu\pi$ is introduced when we express the factorials of negative numbers in terms of the factorials of positive numbers.

3. *Even and odd Legendre functions*

Forming the following functions

$$L_\nu(\cos\theta) = \tfrac{1}{2}[P_\nu(\cos\theta) + P_\nu(-\cos\theta)],$$
$$M_\nu(\cos\theta) = \tfrac{1}{2}[P_\nu(\cos\theta) - P_\nu(-\cos\theta)], \tag{17}$$

we find

$$L_\nu(-x) = L_\nu(x), \qquad M_\nu(-x) = -M_\nu(x),$$
$$L_\nu[\cos(\pi-\theta)] = L_\nu(\cos\theta), \qquad M_\nu[\cos(\pi-\theta)] = -M_\nu(\cos\theta); \tag{18}$$

that is, the L-functions are even functions of x and the M-functions are odd. On account of their symmetry these functions are particularly convenient.

4. *Legendre functions for integral values of $\nu = n$*

If $\nu = n$ is an integer, the series (13) terminates and becomes the *Legendre polynomial of degree n*. If we assume a solution in the form of a series of positive powers of x, and substitute in (3), we obtain

$$\Theta = a_0\left[1 - \frac{n(n+1)}{1\cdot 2}x^2 + \frac{n(n-2)(n+1)(n+3)}{1\cdot 2\cdot 3\cdot 4}x^4 - \cdots\right], \tag{19}$$

if n is even; and

$$\Theta = a_1\left[x - \frac{(n-1)(n+2)}{1\cdot 2\cdot 3}x^3\right.$$
$$\left. + \frac{(n-1)(n-3)(n+2)(n+4)}{1\cdot 2\cdot 3\cdot 4\cdot 5}x^5 - \cdots\right], \tag{20}$$

if n is odd. In order to identify these polynomials with $P_n(x)$ we must determine a_0 and a_1 so that $P_n(1) = 1$. Since (13), (19), (20) are all

polynomials of degree n, we can also make the identification by comparing the coefficients of x^n; these are

$$\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{1 \cdot 2 \cdot 3 \cdots n},$$

$$a_0 \frac{2 \cdot 4 \cdot 6 \cdots n \cdot (n+1)(n+3) \cdots (2n-1)}{1 \cdot 2 \cdot 3 \cdots n} (-)^{n/2}, \tag{21}$$

$$a_1 \frac{2 \cdot 4 \cdot 6 \cdots (n-1) \cdot (n+2)(n+4) \cdots (2n-1)}{1 \cdot 2 \cdot 3 \cdots n} (-)^{(n-1)/2}.$$

Hence,

$$\begin{aligned} a_0 &= (-)^{n/2} \frac{1 \cdot 3 \cdot 5 \cdots (n-1)}{2 \cdot 4 \cdot 6 \cdots n}. \\ a_1 &= (-)^{(n-1)/2} \frac{1 \cdot 3 \cdot 5 \cdots n}{2 \cdot 4 \cdot 6 \cdots (n-1)}. \end{aligned} \tag{22}$$

If we reverse the order of the terms in (19) and (20), we can represent $P_n(x)$ by one expression

$$P_n(x) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{1 \cdot 2 \cdot 3 \cdots n} \left[x^n - \frac{n(n-1)}{2(2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4(2n-1)(2n-3)} x^{n-4} - \cdots \right] \tag{23}$$

This polynomial stops automatically at the right place.

The following special values should be noted:

$$\begin{aligned} &P_n(1) = 1, \qquad P_n(-1) = (-1)^n, \\ &P_{2m+1}(0) = 0, \qquad P_{2m}(0) = (-)^m \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2 \cdot 4 \cdot 6 \cdots 2m}. \end{aligned} \tag{24}$$

Since

$$P_n(-x) = (-)^n P_n(x), \qquad P_n(-\cos\theta) = (-)^n P_n(\cos\theta), \tag{25}$$

we have lost the second independent solution of Legendre's equation. Of the two symmetric solutions (17) one or the other vanishes as ν becomes an integer. The remaining one is equal to P_n $(\cos\theta)$. The second solution may be obtained by differentiation with respect to ν and letting $\nu = n$. Still better we may do as we did in the case of Bessel functions, and introduce another solution which is valid for all nonintegral values

of ν and has a limit as ν approaches an integer. Thus we define the *Legendre function of the second kind* as

$$Q_\nu(x) = \frac{\pi}{2}\,\frac{P_\nu(x)\cos\nu\pi - P_\nu(-x)}{\sin\nu\pi}. \tag{26}$$

As ν approaches an integer, the numerator and denominator approach zero and we have an indeterminate form. Differentiating both terms of the fraction with respect to ν, we obtain

$$Q_n(x) = \tfrac{1}{2}\left[\frac{\partial P_\nu(x)}{\partial\nu} + (-)^{n+1}\frac{\partial P_\nu(-x)}{\partial\nu}\right]_{\nu=n}. \tag{27}$$

Thus for integral values of ν we have the following solutions:

$$\begin{aligned} L_{2m}(x) &= P_{2m}(x) \text{ and } Q_{2m}(x) = \left.\frac{\partial M_\nu(x)}{\partial\nu}\right|_{\nu=2m}, \\ M_{2m+1}(x) &= P_{2m+1}(x) \text{ and } Q_{2m+1}(x) = \left.\frac{\partial L_\nu(x)}{\partial\nu}\right|_{\nu=2m+1}. \end{aligned} \tag{28}$$

Just as L_ν and M_ν form a pair of symmetric solutions for nonintegral values of ν, P_ν and Q_ν form a symmetric pair for integral values of ν. The difference is that L_ν is always an even function and M_ν is always odd; on the other hand, P_n is even and Q_n is odd when n is even while the reverse is true when n is odd.

It can be shown that

$$\begin{aligned} Q_n(\cos\theta) &= P_n(\cos\theta)\log\cot\frac{\theta}{2} - \left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right)P_n(\cos\theta) \\ &+ \sum_{m=1}^{n}\frac{(-)^m(n+m)!}{(m!)^2(n-m)!}\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{m}\right)\sin^{2m}\frac{\theta}{2} \end{aligned} \tag{29}$$

if $n > 0$, and

$$\begin{aligned} &Q_0(\cos\theta) = \log\cot\frac{\theta}{2}, \qquad Q_1(\cos\theta) = \cos\theta\log\cot\frac{\theta}{2} - 1, \\ &Q_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1)\log\cot(\theta/2) - \tfrac{3}{2}\cos\theta. \end{aligned} \tag{30}$$

The above definitions of the Q-functions are for real values of θ and for x in the interval $(-1,1)$. Since all Legendre functions for nonintegral values of ν have a logarithmic singularity either at $x = 1$ or at $x = -1$ and all functions of the second kind are singular at both points, the analytic continuations of these functions to values of x outside $(-1,1)$ will depend on the path of the continuation. What is usually done, however, is to

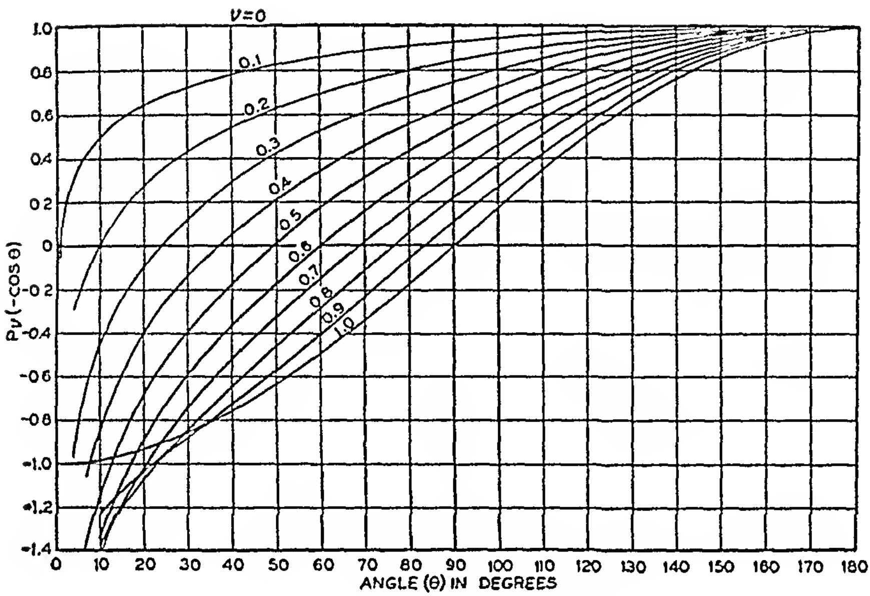

FIG. 21.1. The Legendre function $P_\nu(-\cos\theta)$ for $0 \leq \nu \leq 1$.

define the Q-functions differently for all values of x outside the interval $(-1,1)$. The definition is

$$Q_\nu(x) = \frac{\pi}{2}\,\frac{P_\nu(x)e^{-i\nu\pi} - P_\nu(-x)}{\sin \nu\pi}. \tag{31}$$

For integral values of ν this leads to

$$\begin{aligned} Q_n(x) &= \frac{1}{2}P_n(x)\log\frac{x+1}{x-1} - \left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right)P_n(x) \\ &\quad + \sum_{m=1}^{n}\frac{(-)^m(n+m)!}{(m!)^2(n-m)!}\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{m}\right)\left(\frac{1-x}{2}\right)^m, \end{aligned} \tag{32}$$

for $n > 0$; for $n = 0, 1, 2,$

$$\begin{aligned} Q_0(x) &= \frac{1}{2}\log\frac{x+1}{x-1}, \qquad Q_1(x) = \frac{x}{2}\log\frac{x+1}{x-1} - 1, \\ Q_2(x) &= \frac{3x^2-1}{4}\log\frac{x+1}{x-1} - \frac{3}{2}x. \end{aligned} \tag{33}$$

The difference between (29) and (32) is equal to $\pm\frac{1}{2}i\pi P_n(x)$.

Figures 21.1, 21.2, 21.3, 21.4 illustrate the behavior of Legendre functions for the smaller values of ν and n.

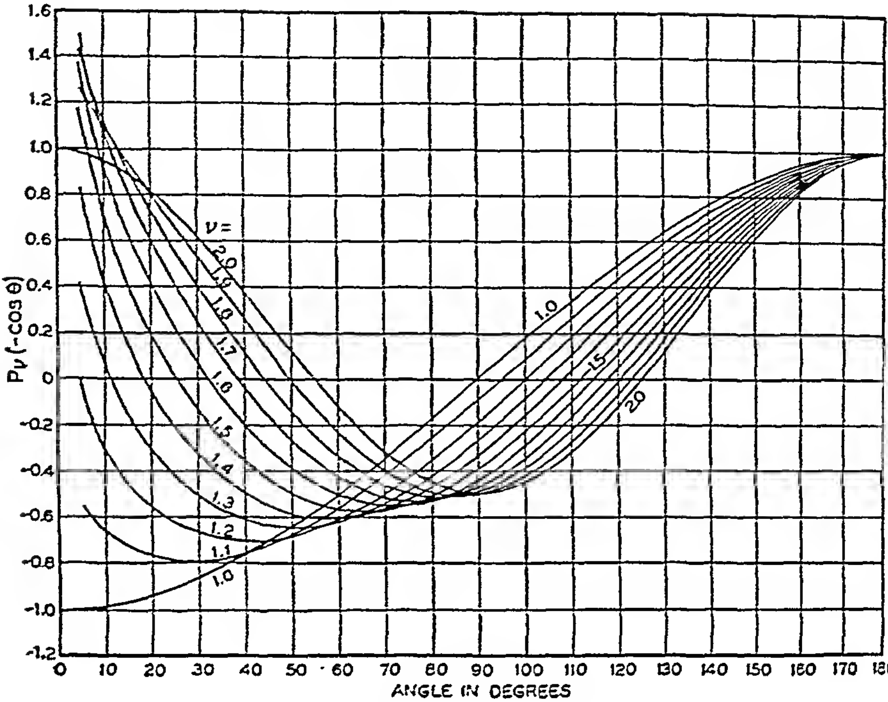

FIG. 21.2. The Legendre function $P_\nu(-\cos\theta)$ for $1 \leq \nu \leq 2$.

5. *Legendre functions and their derivatives*

There is a simple relationship between Legendre functions and their derivatives. Legendre's equation may be written as

$$\frac{d}{d\theta}\left(\sin\theta\,\frac{d\Theta}{d\theta}\right) = -\nu(\nu + 1)\sin\theta\,\Theta. \tag{34}$$

Applying (11–96), we have

$$\Theta_2\,\frac{d\Theta_1}{d\theta} - \Theta_1\,\frac{d\Theta_2}{d\theta} = \frac{C}{\sin\theta}. \tag{35}$$

The constant may be determined for some particular value of θ, $\theta = \pi/2$ for instance. Thus we find

$$Q_\nu(\cos\theta)\,\frac{dP_\nu(\cos\theta)}{d\theta} - P_\nu(\cos\theta)\,\frac{dQ_\nu(\cos\theta)}{d\theta} = \frac{1}{\sin\theta}. \tag{36}$$

Substituting for Q_ν from (26), we have

$$P_\nu(\cos\theta)\,\frac{dP_\nu(-\cos\theta)}{d\theta} - P_\nu(-\cos\theta)\,\frac{dP_\nu(\cos\theta)}{d\theta} = \frac{2\sin\nu\pi}{\pi\sin\theta}. \tag{37}$$

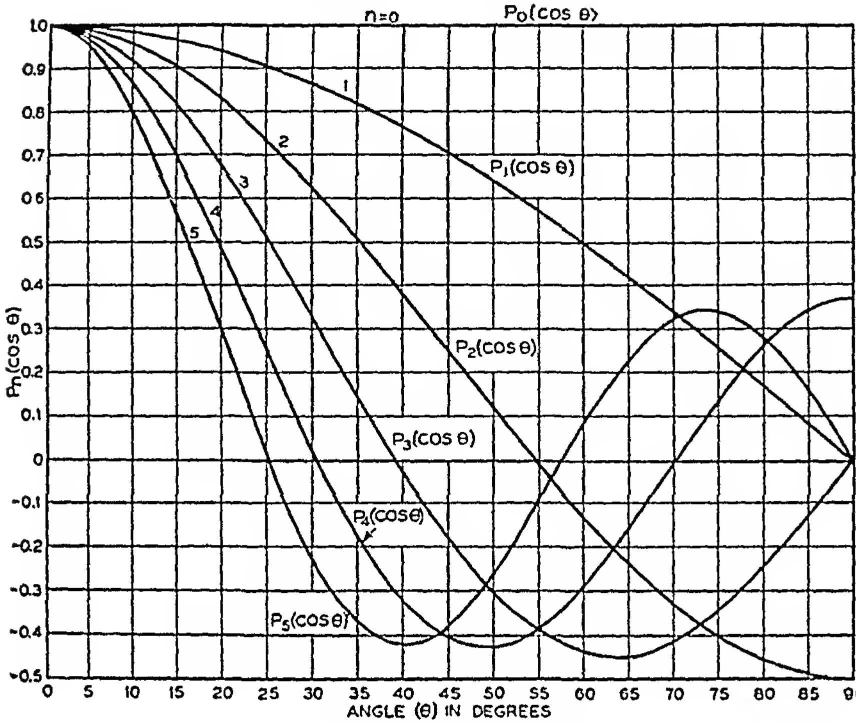

FIG. 21.3. Legendre functions, $P_n(\cos\theta)$, of the first kind for integral values of n.

6. *Integrals of products of Legendre functions*

Starting with the following Legendre's equations

$$\begin{aligned}\frac{d}{d\theta}\left(\sin\theta\,\frac{d\Theta_\nu}{d\theta}\right) &= -\nu(\nu+1)\sin\theta\;\Theta_\nu,\\ \frac{d}{d\theta}\left(\sin\theta\,\frac{d\overline{\Theta}_\mu}{d\theta}\right) &= -\mu(\mu+1)\sin\theta\,\overline{\Theta}_\mu,\end{aligned}\tag{38}$$

multiplying the first by $\overline{\Theta}_\mu$, the second by Θ_ν, subtracting, and integrating, we obtain

$$[\mu(\mu+1)-\nu(\nu+1)]\int\sin\theta\,\Theta_\nu\overline{\Theta}_\mu\,d\theta = \sin\theta\left(\overline{\Theta}_\mu\frac{d\Theta_\nu}{d\theta} - \Theta_\nu\frac{d\overline{\Theta}_\mu}{d\theta}\right). \tag{39}$$

As μ approaches ν, the limiting form of this expression is

$$\int\sin\theta\,\Theta_\nu\overline{\Theta}_\nu\,d\theta = \frac{\sin\theta}{2\nu+1}\left(\frac{\partial\overline{\Theta}_\nu}{\partial\nu}\frac{\partial\Theta_\nu}{\partial\theta} - \Theta_\nu\frac{\partial^2\overline{\Theta}_\nu}{\partial\theta\partial\nu}\right). \tag{40}$$

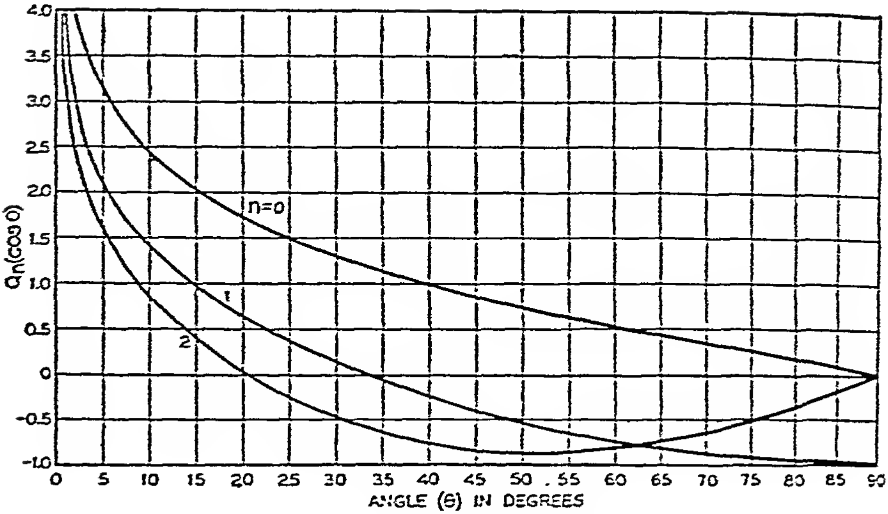

FIG. 21.4. Legendre functions, Q_n (cos θ), of the second kind for integral values of n.

7. *Legendre functions of order $\nu = n + \delta$, where δ is small*

If $\nu = n + \delta$ differs but little from an integer n,

$$P_{n+\delta}(\cos\theta) \simeq P_n(\cos\theta) + \delta \left.\frac{\partial P_\nu}{\partial \nu}\right|_{\nu=n},$$

$$P_{n+\delta}(-\cos\theta) \simeq (-)^n P_n(\cos\theta) + \delta \left.\frac{\partial P_\nu(-\cos\theta)}{\partial \nu}\right|_{\nu=n}. \tag{41}$$

The derivatives of the Legendre functions with respect to ν, when $\nu = n$, are

$$\left.\frac{\partial P_\nu(\cos\theta)}{\partial \nu}\right|_{\nu=n} = 2P_n(\cos\theta)\log\cos(\theta/2) + 2S'_n,$$

$$\frac{\partial P_\nu(-\cos\theta)}{\partial \nu} = 2(-)^n P_n(\cos\theta)\log\sin(\theta/2) + 2S''_n,$$

$$S'_n = \sum_{m=1}^{n} \frac{(-)^m(n+m)!}{m!\,m!\,(n-m)!}\left(\frac{1}{n+m} + \frac{1}{n+m-1} + \cdots + \frac{1}{n+1}\right)\sin^{2m}(\theta/2), \tag{42}$$

$$S''_n = \sum_{m=1}^{n} \frac{(-)^m(n+m)!}{m!\,m!\,(n-m)!}\left(\frac{1}{n+m} + \frac{1}{n+m-1} + \cdots + \frac{1}{n+1}\right)\cos^{2m}(\theta/2).$$

If $n = 0$, $S'_0 = S''_0 = 0$.

We can also treat the problem by the wave perturbation method. Substituting $\nu = n + \delta$ in (38), we have

$$\frac{d}{d\theta}\left(\sin\theta\,\frac{d\Theta_\nu}{d\theta}\right) = -n(n+1)\sin\theta\,\Theta_\nu - \delta(2n+1+\delta)\sin\theta\,\Theta_\nu. \tag{43}$$

Therefore,

$$\Theta_\nu(\theta) = \Theta_n(\theta) - \delta(2n+1+\delta)\int_0^\theta \sin\vartheta\, G_n(\theta,\vartheta)\Theta_\nu(\vartheta)\,d\vartheta, \tag{44}$$

where $G_n(\theta,\vartheta)$ is that solution of the nonperturbed equation ($\delta = 0$) for which

$$G_n(\vartheta,\vartheta) = 0, \qquad \sin\theta\,\frac{dG_n}{d\theta}\bigg|_{\theta=\vartheta} = 1. \tag{45}$$

This solution is

$$G_n(\theta,\vartheta) = P_n(\cos\theta)Q_n(\cos\vartheta) - Q_n(\cos\theta)P_n(\cos\vartheta). \tag{46}$$

In particular,

$$P_\nu(\cos\theta) = P_n(\cos\theta) - \delta(2n+1+\delta)\int_0^\theta \sin\vartheta\, G_n(\theta,\vartheta)P_\nu(\cos\vartheta)\,d\vartheta. \tag{47}$$

To obtain the first perturbation we let $P_\nu(\cos\vartheta) = P_n(\cos\vartheta)$ under the integral sign and apply (40). We can also evaluate the integral directly since G_n is given in terms of elementary functions.

Since the coefficient of δ must be the same as in (41),

$$P_\nu(\cos\theta) \cong P_n(\cos\theta) - \delta\left(1 + \frac{\delta}{2n+1}\right)\frac{\partial P_\nu}{\partial\nu}\bigg|_{\nu=n}\cdot \tag{48}$$

In particular

$$P_\delta(\cos\theta) = 1 + 2\delta(1+\delta)\log\cos(\theta/2). \tag{49}$$

8. *Associated Legendre functions*

When the variables in the Laplace equation or the wave equation expressed in spherical coordinates are separated, we obtain

$$\frac{d}{d\theta}\left\{\sin\theta\,\frac{d\Theta}{d\theta}\right\} = \left[-\nu(\nu+1)\sin\theta + \frac{\mu^2}{\sin\theta}\right]\Theta. \tag{50}$$

This is the *associated Legendre equation* and its solutions are *associated Legendre functions.*

When θ is small, (50) is approximated by the Bessel equation of order μ, thus for small θ we have

$$\Theta(\theta) \simeq AJ_\mu(\beta\theta) + BN_\mu(\beta\theta), \qquad \beta = \sqrt{\nu(\nu+1)}. \tag{51}$$

Removing the term containing the first derivative by means of the transformation (4), we have

$$\frac{d^2\hat{\Theta}}{d\theta^2} = -[(\nu + \tfrac{1}{2})^2 + \tfrac{1}{4} - \mu^2 - (\mu^2 - \tfrac{1}{4})\cot^2\theta]\hat{\Theta}. \tag{52}$$

In the vicinity of $\theta = \pi/2$ the last bracketed term is negligible and

$$\hat{\Theta} = A'\cos k\theta + B'\sin k\theta, \qquad k = \sqrt{(\nu + \tfrac{1}{2})^2 + \tfrac{1}{4} - \mu^2}. \tag{53}$$

If θ is imaginary, $\theta = iu$, and large, then $\cot\theta = -i\coth u \simeq -i$; equation (52) becomes approximately

$$\frac{d^2\hat{\Theta}}{du^2} = (\nu + \tfrac{1}{2})^2\hat{\Theta}, \tag{54}$$

and its solutions are

$$\begin{aligned}\hat{\Theta}(u) &= A''\cosh(\nu + \tfrac{1}{2})u + B''\sinh(\nu + \tfrac{1}{2})u \\ &= C\exp(\nu + \tfrac{1}{2})u + D\exp[-(\nu + \tfrac{1}{2})u].\end{aligned} \tag{55}$$

These approximations are also valid for complex u when the real part is large.

For general values of μ and ν, and complex values of x such that $|1 - x| < 2$, the standard form of solution of the associated Legendre equation is usually taken to be

$$P_\nu^\mu(x) = \frac{(\mu - 1)!\sin\mu\pi}{\pi}\left(\frac{x+1}{x-1}\right)^{\mu/2} F\left(-\nu, \nu + 1; 1 - \mu; \frac{1-x}{2}\right). \tag{56}$$

If μ is a positive integer m, this formula has to be modified, and the general form is

$$P_\nu^m(x) = \frac{(-1)^m(\nu + m)!}{2^m(\nu - m)!m!}(x^2 - 1)^{m/2}F\left(m - \nu, m + \nu + 1; m + 1; \frac{1-x}{2}\right). \tag{57}$$

If ν is a positive integer n, greater than m, the hypergeometric function is a polynomial of degree $n - m$.

For real values of x in the interval $(-1,+1)$, that is, for real θ, it is customary to define the associated Legendre function as

$$P_\nu^\mu(x) = e^{i\mu\pi/2}P_\nu^\mu(x + i\cdot 0) = e^{-i\mu\pi/2}P_\nu^\mu(x - i\cdot 0), \tag{58}$$

where the functions on the right have their usual meanings for complex values of x. This makes $P_\nu^\mu(\cos\theta)$ real for real values of θ, and for integral

values of μ (57) becomes

$$P_\nu^m(\cos\theta) = \frac{(-1)^m(\nu+m)!\sin^m\theta}{2^m(\nu-m)!\,m!}F\left(m-\nu, m+\nu+1; m+1; \sin^2\frac{\theta}{2}\right). \quad (57')$$

The series (56) and (57) converge only if $|1-x|<2$; if $|x|>1$, then a general series for P_ν^μ, valid except when ν is half an odd integer, is

$$P_\nu^\mu(x) = \frac{\sin(\nu+\mu)\pi}{2^{\nu+1}\cos\nu\pi}\frac{(\nu+\mu)!}{(\nu+\frac{1}{2})!(-\frac{1}{2})!}(x^2-1)^{\mu/2}x^{-\nu-\mu-1}$$
$$\times F\left(\frac{\mu+\nu+2}{2}, \frac{\mu+\nu+1}{2}; \nu+\frac{3}{2}; x^{-2}\right) \quad (59)$$
$$+\frac{2^\nu(\nu-\frac{1}{2})!}{(\nu-\mu)!(-\frac{1}{2})!}(x^2-1)^{\mu/2}x^{\nu-\mu}F\left(\frac{\mu-\nu+1}{2}, \frac{\mu-\nu}{2}; \frac{1}{2}-\nu; x^{-2}\right).$$

There are many other series which may be found in E. W. Hobson's treatise on *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press).

An *associated Legendre function of the second kind* for real θ is defined as follows

$$Q_\nu^\mu(\cos\theta) = \frac{\pi}{2}\frac{P_\nu^\mu(\cos\theta)\cos(\nu+\mu)\pi - P_\nu^\mu(-\cos\theta)}{\sin(\nu+\mu)\pi}. \quad (60)$$

Everywhere else in the complex plane the definition is

$$Q_\nu^\mu(x) = \frac{\pi}{2}e^{i\mu\pi}\frac{P_\nu^\mu(x)e^{\mp i\nu\pi} - P_\nu^\mu(-x)}{\sin(\nu+\mu)\pi}. \quad (61)$$

The upper sign should be taken if $\mathrm{re}(x)>0$, and the lower if $\mathrm{re}(x)<0$.

Normally $P_\nu^\mu(x)$ and $P_\nu^\mu(-x)$ are linearly independent solutions of Legendre's equation; but when $\nu+\mu$ is an integer, they are not. In this exceptional case, the limit of Q_ν^μ exists and may be chosen as the second solution.

T. M. MacRobert in his treatise on *Spherical Harmonics* (E. P. Dutton and Company) employs a function $T_n^m(x)$ for integral values of m and n and for $-1<x<1$. This function is identical with $(-)^mP_n^m(x)$ as defined in this section.

It should be kept in mind that if one is concerned with a limited range of the independent variable in which approximate solutions of Legendre's equation are available, it is unnecessary to employ standard Legendre functions. The latter are needed only if we are interested in a range where no one approximation is valid; even then it may be found more

convenient to join the approximations in the various parts of the range instead of employing the standard forms.

9. *Orthogonal expansions*

In Section 13.10 we have seen that solutions of the wave equation in spherical coordinates are of the form $R(r)\Theta(\theta)\Phi(\varphi)$, where Θ is an associated Legendre function and $\Phi(\varphi) = A \cos \mu\varphi + B \sin \mu\varphi$. The products $\Theta(\theta)\Phi(\varphi)$ are known as *spherical harmonics*.

The parameters μ and ν are determined by the boundary conditions. For instance, for waves in homogeneous space Φ must be a periodic function and this requires μ to be an integer. Along the rays $\theta = 0, \pi$ the function Θ must be finite; this requires ν to be an integer and restricts Θ to being proportional to $P_n^m(\cos \theta)$.

In problems involving "tesseral horns," bounded by two cones $\theta = \theta_1$ and $\theta = \theta_2$ and two half-planes $\varphi = \varphi_1$ and $\varphi = \varphi_2$, the wave function or its normal derivative may be required to vanish or be subjected to some other boundary condition. If $\Theta(\theta)\Phi(\varphi)$ is to vanish on the half-planes $\varphi = \varphi_1$ and $\varphi = \varphi_2$, then $\Phi(\varphi) = A \sin \mu(\varphi - \varphi_1)$ and $\mu = m\pi/(\varphi_2 - \varphi_1)$, where m is an integer. To make the function equal to zero on $\theta = \theta_1$ and $\theta = \theta_2$, we have to set

$$\Theta(\theta) = P_\nu^\mu(-\cos \theta_1)P_\nu^\mu(\cos \theta) - P_\nu^\mu(\cos \theta_1)P_\nu^\mu(-\cos \theta), \tag{62}$$

where ν is a root of the characteristic equation

$$\frac{P_\nu^\mu(\cos \theta_1)}{P_\nu^\mu(-\cos \theta_1)} = \frac{P_\nu^\mu(\cos \theta_2)}{P_\nu^\mu(-\cos \theta_2)}. \tag{63}$$

On the surface of some sphere $r = a$ the wave function may be required to reduce to a given function $F(\theta, \varphi)$. If this function does not coincide with any of the spherical harmonics, we face the problem of expanding the function into appropriate harmonics so that the values on $r = a$ can be analytically extended to other values of r, while preserving the specified boundary conditions. The calculation of the coefficients is greatly facilitated by the fact that, for many boundary conditions, spherical harmonics are orthogonal.

The Φ-functions belonging to different characteristic values of μ are obviously orthogonal; and we have only to show that the associated Legendre functions belonging to the same μ but different characteristic values of ν are orthogonal. Starting with

$$\begin{aligned} \frac{d}{d\theta}\left(\sin \theta \frac{d\Theta}{d\theta}\right) &= \left[-\nu(\nu + 1) \sin \theta + \frac{\mu^2}{\sin \theta}\right] \Theta, \\ \frac{d}{d\theta}\left(\sin \theta \frac{d\bar{\Theta}}{d\partial}\right) &= \left[-\bar{\nu}(\bar{\nu} + 1) \sin \theta + \frac{\mu^2}{\sin \theta}\right] \bar{\Theta}, \end{aligned} \tag{64}$$

multiplying the first by $\bar{\Theta}$ and the second by Θ, subtracting, and integrating from $\theta = \theta_1$ to $\theta = \theta_2$, we obtain

$$[\bar{\nu}(\bar{\nu}+1)-\nu(\nu+1)]\int_{\theta_1}^{\theta_2} \sin\theta\,\Theta(\theta)\bar{\Theta}(\theta)d\theta = \sin\theta\left(\bar{\Theta}\frac{d\Theta}{d\theta} - \Theta\frac{d\bar{\Theta}}{d\theta}\right)\Bigg|_{\theta_1}^{\theta_2}. \quad (65)$$

The Θ's are orthogonal when the function on the right vanishes at each limit of integration. This happens if ν and $\bar{\nu}$ are distinct roots of (63). More generally the Θ's are orthogonal if ν and $\bar{\nu}$ are distinct roots of

$$\sin\Theta\frac{d\Theta}{d\theta} + k\Theta = 0, \quad (66)$$

where k is an arbitrary constant. This constant may be different for the two limits. This generalization is obtained by the simple device of adding and subtracting $k\Theta\bar{\Theta}$ on the right in (65).

The derivatives of Legendre functions are the associated Legendre functions with $\mu = 1$; hence, they also form an orthogonal set.

10. *Miscellaneous formulas*

The following are some of the more frequently useful formulas in addition to those given in the preceding sections:

$$P_0(\cos\theta) = 1, \qquad P_1(\cos\theta) = \cos\theta, \qquad P_1^1(\cos\theta) = -\sin\theta.$$

$$P_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1) = \tfrac{1}{4}(3\cos 2\theta + 1),$$

$$P_2^1(\cos\theta) = -3\sin\theta\cos\theta, \qquad P_2^2(\cos\theta) = 3\sin^2\theta,$$

$$P_3(\cos\theta) = \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta) = \tfrac{1}{8}(5\cos 3\theta + 3\cos\theta),$$

$$P_3^1(\cos\theta) = -\tfrac{3}{2}\sin\theta(5\cos^2\theta - 1), \qquad P_3^2 = 15\sin^2\theta\cos\theta, \quad (67)$$

$$P_3^3 = -15\sin^3\theta,$$

$$P_4 = \tfrac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3) = \tfrac{1}{64}(35\cos 4\theta + 20\cos 2\theta + 9),$$

$$P_4^1 = -\tfrac{5}{2}\sin\theta(7\cos^3\theta - 3\cos\theta), \qquad P_4^2 = \tfrac{15}{2}\sin^2\theta(7\cos^2\theta - 1),$$

$$P_4^3 = -105\sin^3\theta\cos\theta, \qquad P_4^4 = 105\sin^4\theta.$$

$$P_n^1(\cos\theta) = \frac{dP_n(\cos\theta)}{d\theta}, \qquad P_n^m(\cos\theta) = (-)^m\sin^m\theta\frac{d^mP_n}{d(\cos\theta)^m}. \quad (68)$$

$$P_\nu^\mu(\cos\theta) = e^{i\mu\pi/2}P_\nu^\mu(\cos\theta + 0\cdot i) = e^{-i\mu\pi/2}P_\nu^\mu(\cos\theta - 0\cdot i). \quad (69)$$

$$e^{i\mu\pi}Q_\nu^\mu(\cos\theta) = \tfrac{1}{2}[e^{-i\mu\pi/2}Q_\nu^\mu(\cos\theta + 0\cdot i) + e^{i\mu\pi/2}Q_\nu^\mu(\cos\theta - 0\cdot i)]. \quad (70)$$

$$P_\nu(\cos\theta) = \frac{\sin\nu\pi}{\pi}\sum_{s=0}^{\infty}\frac{(-)^s(\nu+s)!}{(\nu-s)!}\,[2\log\cos(\theta/2) + \Psi(\nu+s)$$

$$+\ \Psi(\nu-s) - 2\Psi(s)]\,\frac{\cos^{2s}(\theta/2)}{(s!)^2} + \cos\nu\pi\sum_{s=0}^{\infty}\frac{(-)^s(\nu+s)!}{(\nu-s)!(s!)^2}\cos^{2s}(\theta/2). \tag{71}$$

(This expansion is particularly useful in the vicinity of $\theta = \pi$; see equation (13) for an expansion in the vicinity of $\theta = 0$.)

$$P_\nu^\mu(\cos\theta)$$

$$= 2^\mu\cos\frac{(\nu+\mu)\pi}{2}\,\frac{\left(\frac{\nu+\mu-1}{2}\right)!}{\left(\frac{\nu-\mu}{2}\right)!\left(-\frac{1}{2}\right)!}\sin^\mu\theta\,F\left(\frac{\nu+\mu+1}{2}, \frac{-\nu+\mu}{2}; \frac{1}{2}; \cos^2\theta\right)$$

$$+\ 2^{\mu+1}\sin\frac{(\nu+\mu)\pi}{2}\,\frac{\left(\frac{\nu+\mu}{2}\right)!}{\left(\frac{\nu-\mu-1}{2}\right)!\left(-\frac{1}{2}\right)!}\cos\theta\sin^\mu\theta \tag{72}$$

$$\times\ F\left(\frac{\nu+\mu+2}{2}, \frac{-\nu+\mu+1}{2}; \frac{3}{2}; \cos^2\theta\right).$$

(This expansion is particularly useful in the vicinity of $\theta = \pi/2$.)

$$P_\nu^\mu(0) = 2^\mu\cos\frac{(\nu+\mu)\pi}{2}\,\frac{\left(\frac{\nu+\mu-1}{2}\right)!}{\left(\frac{\nu-\mu}{2}\right)!\left(-\frac{1}{2}\right)!}. \tag{73}$$

$$\frac{d}{d\theta}P_\nu^\mu(\cos\theta)\Big|_{\theta=\pi/2} = -2^{\mu+1}\sin\frac{(\nu+\mu)\pi}{2}\,\frac{\left(\frac{\nu+\mu}{2}\right)!}{\left(\frac{\nu-\mu-1}{2}\right)!\left(-\frac{1}{2}\right)!}. \tag{74}$$

$$Q_\nu^\mu(0) = -2^{\mu-1}\sin\frac{(\nu+\mu)\pi}{2}\,\frac{\left(\frac{\nu+\mu-1}{2}\right)!\left(-\frac{1}{2}\right)!}{\left(\frac{\nu-\mu}{2}\right)!}. \tag{75}$$

$$\frac{d}{d\theta} Q_\nu^\mu (\cos\theta)\bigg|_{\theta=\pi/2} = -2^\mu \cos\frac{(\nu+\mu)\pi}{2} \frac{\left(\frac{\nu+\mu}{2}\right)!\left(-\frac{1}{2}\right)!}{\left(\frac{\nu-\mu-1}{2}\right)!}. \tag{76}$$

$$\begin{aligned} &-P_\nu^\mu (\cos\theta)\frac{d}{d\theta} Q_\nu^\mu (\cos\theta) + Q_\nu^\mu (\cos\theta)\frac{d}{d\theta} P_\nu^\mu (\cos\theta) \\ &= \frac{2^{2\mu}\left(\frac{\nu+\mu-1}{2}\right)!\left(\frac{\nu+\mu}{2}\right)!}{\left(\frac{\nu-\mu-1}{2}\right)!\left(\frac{\nu-\mu}{2}\right)!\sin\theta} = \frac{(\nu+\mu)!}{(\nu-\mu)!\sin\theta}. \end{aligned} \tag{77}$$

$$\begin{aligned} &P_\nu^\mu (\cos\theta)\frac{d}{d\theta} P_\nu^\mu (-\cos\theta) - P_\nu^\mu (-\cos\theta)\frac{d}{d\theta} P_\nu^\mu (\cos\theta) \\ &= \frac{2^{2\mu+1}\left(\frac{\nu+\mu-1}{2}\right)!\left(\frac{\nu+\mu}{2}\right)!\sin(\nu+\mu)\pi}{\pi\left(\frac{\nu-\mu-1}{2}\right)!\left(\frac{\nu-\mu}{2}\right)!\sin\theta} = \frac{2(\nu+\mu)!\sin(\nu+\mu)\pi}{\pi(\nu-\mu)!\sin\theta}. \end{aligned} \tag{78}$$

$$(2\nu+1)xP_\nu^\mu(x) - (\nu-\mu+1)P_{\nu+1}^\mu(x) - (\nu+\mu)P_{\nu-1}^\mu(x) = 0. \tag{79}$$

$$(x^2-1)\frac{d}{dx} P_\nu^\mu(x) = (\nu-\mu+1)P_{\nu+1}^\mu(x) - (\nu+1)xP_\nu^\mu(x) \tag{80}$$

$$(x^2-1)\frac{d}{dx} P_\nu^\mu(x) = \nu xP_\nu^\mu(x) - (\nu+\mu)P_{\nu-1}^\mu(x). \tag{81}$$

$$P_\nu^{\mu+2}(x)+2(\mu+1)x(x^2-1)^{-1/2}P_\nu^{\mu+1}(x)-(\nu-\mu)(\nu+\mu+1)P_\nu^\mu(x)=0. \tag{82}$$

$$P_\nu^{\mu+2}(\cos\theta)+2(\mu+1)\cot\theta P_\nu^{\mu+1}(\cos\theta)+(\nu-\mu)(\nu+\mu+1)P_\nu^\mu(\cos\theta)=0. \tag{83}$$

$$\begin{aligned} F(\theta,\varphi) &= \sum_{n=0}^\infty a_n P_n(\cos\theta) \\ &\quad + \sum_{n=1}^\infty \sum_{m=1}^\infty (a_{n,m}\cos m\varphi + b_{n,m}\sin m\varphi)P_n^m(\cos\theta), \\ a_n &= \frac{2n+1}{4\pi}\int_0^{2\pi}\int_0^\pi \sin\hat\theta\, P_n(\cos\hat\theta)F(\hat\theta,\hat\varphi)\,d\hat\theta\,d\hat\varphi, \\ a_{n,m} &= \frac{2n+1}{2\pi}\frac{(n-m)!}{(n+m)!}\int_0^{2\pi}\int_0^\pi \sin\hat\theta P_n^m(\cos\hat\theta)\cos m\hat\varphi F(\hat\theta,\hat\varphi)\,d\hat\theta\,d\hat\varphi, \\ b_{n,m} &= \frac{2n+1}{2\pi}\frac{(n-m)!}{(n+m)!}\int_0^{2\pi}\int_0^\pi \sin\hat\theta P_n^m(\cos\hat\theta)\sin m\hat\varphi F(\hat\theta,\hat\varphi)\,d\hat\theta\,d\hat\varphi. \end{aligned} \tag{84}$$

$$e^{-i\beta r\cos\theta} = \frac{1}{\beta r}\sum_{n=0}^{\infty}(2n+1)(-i)^n \hat{J}_n(\beta r)P_n(\cos\theta). \tag{85}$$

$$e^{i\beta r\cos\theta} = \frac{1}{\beta r}\sum_{n=0}^{\infty}(2n+1)i^n \hat{J}_n(\beta r)P_n(\cos\theta). \tag{86}$$

$$\begin{aligned}\frac{e^{-\sigma R}}{R} &= \frac{1}{\Gamma a r}\sum_{n=0}^{\infty}(2n+1)\hat{K}_n(\sigma a)\hat{I}_n(\sigma r)P_n(\cos\theta), \qquad r < a,\\ &= \frac{1}{\Gamma a r}\sum_{n=0}^{\infty}(2n+1)\hat{I}_n(\sigma a)\hat{K}_n(\sigma r)P_n(\cos\theta), \qquad r > a.\\ &\qquad R = (r^2 + a^2 - 2ar\cos\theta)^{1/2}.\end{aligned} \tag{87}$$

$$\begin{aligned}\frac{e^{-i\beta R}}{R} &= \frac{1}{i\beta a r}\sum_{n=0}^{\infty}(2n+1)[\hat{J}_n(\beta a)-i\hat{N}_n(\beta a)]\hat{J}_n(\beta r)P_n(\cos\theta),\ r < a,\\ &= \frac{1}{i\beta a r}\sum_{n=0}^{\infty}(2n+1)\hat{J}_n(\beta a)[\hat{J}_n(\beta r)-i\hat{N}_n(\beta r)]P_n(\cos\theta), \quad r > a.\end{aligned} \tag{88}$$

$$\begin{aligned}&P_n[\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]\\ &\quad = P_n(\cos\theta_1)P_n(\cos\theta_2)\\ &\quad + 2\sum_{m=1}^{m=n}\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta_1)P_n^m(\cos\theta_2)\cos m(\varphi_1-\varphi_2).\end{aligned} \tag{89}$$

$$\begin{aligned}\int_0^{\pi}\sin\theta[P_n(\cos\theta)]^2\,d\theta &= \frac{2}{2n+1},\\ \int_0^{\pi}\sin\theta[P_n^m(\cos\theta)]^2\,d\theta &= \frac{2}{2n+1}\,\frac{(n+m)!}{(n-m)!}.\end{aligned} \tag{90}$$

$$\int_{\theta_1}^{\theta_2}\sin\theta\,\Theta_\nu(\theta)\,\overline{\Theta}_\nu(\theta)\,d\theta = \frac{\sin\theta}{2\nu+1}\left[\frac{\partial\overline{\Theta}_\nu}{\partial\nu}\,\frac{\partial\Theta_\nu}{\partial\theta} - \Theta_\nu\,\frac{\partial^2\overline{\Theta}_\nu}{\partial\theta\,\partial\nu}\right]_{\theta_1}^{\theta_2}. \tag{91}$$

(see equation 65)

As $\theta \to 0$, $\quad P_\nu(\cos\theta) \to J_0(\beta\theta) - \frac{1}{12}\theta^2 J_2(\beta\theta), \qquad \beta^2 = \nu(\nu+1).$ (92)

CHAPTER XXII

Formulation of Equations

Knowledge of physical laws, understanding of mathematical concepts, and good judgment, are required for success in translating a given physical problem into mathematical equations. Good judgment is needed to sift out unimportant factors and make simplifying assumptions without impairing the usefulness of the results. " Formulation of equations " is a field of endeavor in which science and mathematics meet. In it there is more science than mathematics; mathematical concepts are involved, but not mathematical methods.

It is only through continued study of the various branches of theoretical physics — mechanics, elasticity, electromagnetic theory, etc. — that the student gradually acquires the knowledge and experience which enable him to express a given physical problem in the appropriate mathematical form. As his experience increases he attains a better understanding of mathematical ideas and methods, and this understanding in turn leads to a deeper insight into natural phenomena.

In this chapter we shall give some illustrations of the equations which arise in various types of physical problems. We shall find that some very general mathematical equations are applicable to problems in different fields of physics, and that frequently the main task of the applied mathematician is to determine the most convenient form of solution of these equations, when the actual physical conditions inherent in a given problem are taken into consideration.

1. *Motion of a projectile in vacuum*

The center of gravity of a projectile moves as if the entire mass were concentrated at this point. Its motion is governed by the first and second laws of Newton, Reference 1, p. 54; R-2, pp. 50 and 92; and R-3, p. 248. Let the x and y-axes be horizontal, at the surface of the earth; then, for mass m at point $P(x,y,z)$ we have

$$\frac{dx}{dt} = \text{constant}, \quad \frac{dy}{dt} = \text{constant}, \quad m\frac{d^2z}{dt^2} = -mg,$$

where g is the acceleration of gravity and mg is the downward force of gravity. The constants of integration are determined by the initial posi-

tion and the initial velocity of the particle. Complete the solution and compare with p. 64, R-1; or p. 94, R-2; or p. 251, R-3. Find the horizontal range of the projectile.

2. *Motion of an electron in crossed uniform electric and magnetic fields*

To solve this problem we need to know the expressions for the force on a charged particle in an electromagnetic field in addition to Newton's laws of motion. From pp. 61 and 72 of R-4, the total force on a charge q is

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B},$$

where $\vec{E}$ is the electric intensity, $\vec{B}$ is the magnetic displacement density, and $\vec{v}$ is the velocity of the particle. Let $\vec{E}$ and $\vec{B}$ be respectively parallel to the x and y-axes; then

$$\vec{E} = iE, \qquad \vec{B} = jB, \qquad \vec{v} = i\frac{dx}{dt} + j\frac{dy}{dt} + k\frac{dz}{dt}$$

$$\vec{v} \times \vec{B} = i \times jB\frac{dx}{dt} + k \times jB\frac{dz}{dt} = kB\frac{dx}{dt} - iB\frac{dz}{dt}.$$

If m is the mass of the charged particle, by Newton's law

$$m\frac{d^2x}{dt^2} = qE - qB\frac{dz}{dt}, \qquad m\frac{d^2y}{dt^2} = 0, \qquad m\frac{d^2z}{dt^2} = qB\frac{dx}{dt}.$$

Solve the problem for the case in which at $t = 0$ the particle is at rest at the origin.

Ans. $x = \dfrac{E}{\omega B}(1 - \cos \omega t), \qquad y = 0, \qquad z = \dfrac{E}{\omega B}(\omega t - \sin \omega t),$

$\omega = qB/m.$

3. *Radioactivity*

Some substances disintegrate spontaneously. A reasonable assumption is that the rate of disintegration per unit mass of a given radioactive substance is constant,

$$\frac{1}{m}\frac{dm}{dt} = -k.$$

Hence, $m(t) = m(0) \exp(-kt)$ and the assumption can be verified experimentally. Express the "half-life" of the substance in terms of k.

Ans. $(\log 2)/k = 0.693/k.$

4. *Chemical reactions*

Chemical reactions obey " mass laws " similar to that in the preceding section. If the substance x decomposes and forms two substances y and z,

$$\frac{dx}{dt} = -k_1x, \qquad \frac{dy}{dt} = k_2x, \qquad \frac{dz}{dt} = k_3x;$$

that is, a given fraction of x decomposes in each unit of time, and y and z are formed in proportion to the amount of x present. The integration of these equations introduces three arbitrary constants in addition to k_1, k_2, k_3. If the amounts of x, y, z are determined experimentally at two different instants, there will be enough equations to calculate all these constants.

5. *Consecutive unimolecular reactions*

If the substance x decomposes into y and y into z, then the rate of decomposition of x is proportional to x and the rate of formation of z is proportional to y. The rate of change of y must be the difference of the rates of its formation on the one hand and its decomposition on the other. Thus

$$\frac{dx}{dt} = -k_1x, \qquad \frac{dz}{dt} = k_2y, \qquad \frac{dy}{dt} = k_1x - k_2y.$$

The first equation may be solved first and the result substituted in the last equation; then the last equation may be solved and the result substituted in the second.

6. *Consecutive bimolecular reactions*

Let a and b be the initial molar concentrations of two reacting substances, and let x be the amount transformed in the interval of time t; on p. 429 of R-5 we find the following equation for the velocity of the reaction:

$$\frac{dx}{dt} = k(a - x)(b - x).$$

Solving for kt, we have

$$kt = \int_0^x \frac{dx}{(a - x)(b - x)}.$$

To integrate, express the integrand in terms of partial fractions

$$\frac{1}{(a - x)(b - x)} = \frac{A}{a - x} + \frac{B}{b - x}.$$

To find A and B, multiply by $(a - x)(b - x)$ and then let x assume the values a and b in succession.

$$\textit{Ans.}\quad kt = \frac{1}{a-b}\log\frac{b(a-x)}{a(b-x)} \quad\text{or}\quad \frac{b(a-x)}{a(b-x)} = e^{k(a-b)t}.$$

7. *Simple pendulum*

In a "simple pendulum," Figure 22.1, the entire mass M is supposed to be concentrated at the end. The force of gravity is Mg, where g is the acceleration of gravity; but only its component, $Mg \sin\theta$, is effective in altering the angular coordinate θ — it tends to decrease it. By Newton's second law (see the references in Problem 1):

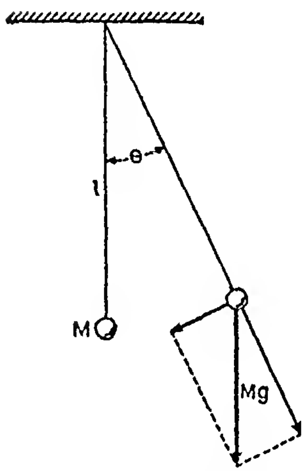

FIG. 22.1. A simple pendulum.

$$M\frac{d^2(l\theta)}{dt^2} = -Mg\sin\theta, \qquad \frac{d^2\theta}{dt^2} = -\frac{g}{l}\sin\theta.$$

The exact solution of this equation may be expressed in terms of elliptic integrals (see Problem 5 of Chapter 4, Section 5). For small angles $\sin\theta \cong \theta$ and the solution is sinusoidal.

Solve the above equation and compare the solution with p. 85 of R-1, p. 97 of R-2 and with pp. 310–314 of R-3.

8. *Conservation of energy*

The law of conservation of energy may conveniently be used to obtain the equation of motion of a system of one degree of freedom. If the system is nondissipative, the total energy $\mathcal{E}$ is the sum of the kinetic energy T and the potential energy V,

$$\mathcal{E} = T + V = \text{constant}. \tag{1}$$

If the system is dissipative,

$$\frac{d\mathcal{E}}{dt} = \frac{d}{dt}(T + V) = -P, \tag{2}$$

where P is the rate of heat generation.

In Problem 7 the velocity of M is $v = l\dot{\theta}$, where $\dot{\theta}$ is the time derivative of θ, and the kinetic energy is

$$T = \tfrac{1}{2}Mv^2 = \tfrac{1}{2}Ml^2\dot{\theta}^2.$$

The potential energy is the product of the force of gravity Mg and the height of M above ground; thus

$$V = Mg\ell(1 - \cos\theta) + \text{constant}.$$

The first term represents the acquisition of potential energy as M rises from its lowest position to that given by the angle θ; the second term does not affect the motion of the pendulum. By (1)

$$\tfrac{1}{2}M\ell^2\dot{\theta}^2 + Mg\ell(1 - \cos\theta) = \mathcal{E},$$

where $\mathcal{E}$ is now the kinetic energy when $\theta = 0$ or the potential energy when $\dot{\theta} = 0$. If we differentiate this equation, we shall obtain the equation given in the preceding problem. We may also solve the above equation for $\dot{\theta} = d\theta/dt$, then for dt, and integrate. The integral so obtained is called the elliptic integral of the first kind; it reduces to the inverse sine when θ is small.

In the case of free oscillations of a mass M attached to a spring with stiffness S, Figure 1.16, equation (1) becomes

$$\tfrac{1}{2}M\dot{x}^2 + \tfrac{1}{2}Sx^2 = \text{constant},$$

where x is the displacement of M from the neutral position. Differentiating with respect to t, we obtain the usual form of the equation for free oscillations.

For an electric circuit containing an inductance L, a capacitance C, and a resistance R, equation (2) becomes

$$\frac{d}{dt}\left(\frac{1}{2}L\dot{q}^2 + \frac{1}{2}\frac{q^2}{C}\right) = -R\dot{q}^2.$$

In this equation, q is the charge on the plates of the capacitor and $\dot{q} = dq/dt$ is the current in the circuit.

9. *Compound pendulum*

The problems in the preceding section are too simple to show the advantages of the conservation of energy principle for obtaining equations of motion. We shall now consider a more difficult problem in which we cannot use Newton's laws directly. In a "compound pendulum" the mass of the rod is not zero; let it be m. If it is uniformly distributed and if x is the distance from the point of support,

$$T = \tfrac{1}{2}M\ell^2\dot{\theta}^2 + \int_0^\ell \frac{1}{2}\frac{m}{\ell}x^2\dot{\theta}^2\,dx = \tfrac{1}{2}(M + \tfrac{1}{3}m)\ell^2\dot{\theta}^2,$$

$$V = Mg\ell(1 - \cos\theta) + \int_0^\ell \frac{m}{\ell}g(1 - \cos\theta)x\,dx = (M + \tfrac{1}{2}m)g\ell(1 - \cos\theta).$$

If the mass of the rod is not uniformly distributed, we must replace m/l by the linear density $\delta(x)$.

10. *Systems with several degrees of freedom*

Consider a particle of mass m, moving under the influence of the gravitational force exerted by a mass M at a fixed point O. Let the plane defined by O and the initial velocity of m be the xy-plane. By this choice, the initial velocity and the force in the z direction are made equal to zero; hence the particle will remain in the xy-plane, its position may be given by two coordinates, and we have a moving system with two degrees of freedom. To obtain the equations of motion, we combine Newton's second law of motion (R-3, p. 248) with Newton's law of gravitation (R-3, p. 38); thus the force of gravitation is $GMm\rho^{-2}$, where ρ is the distance between M and m, G is the gravitational constant, 6.66×10^{-11} newtons or kilogram-meters per second per second, and the equations of motion are, Figure 22.2,

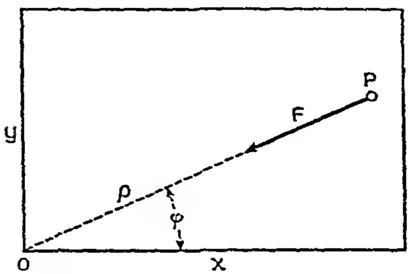

FIG. 22.2. A force F acting toward a fixed point O.

$$\ddot{x} = -GM\rho^{-2}\cos\varphi = -GMx(x^2 + y^2)^{-3/2},$$

$$\ddot{y} = -GM\rho^{-2}\sin\varphi = -GMy(x^2 + y^2)^{-3/2},$$

where the dots indicate differentiation with respect to time t.

The above equations of motion become simpler if expressed in terms of polar coordinates (ρ,φ). The direct transformation of $\ddot{x}$ and $\ddot{y}$ from cartesian to polar coordinates is given in R-3, p. 237; using these results we have

$$\ddot{\rho} - \rho\dot{\varphi}^2 = -GM\rho^{-2}, \qquad \frac{1}{\rho}\frac{d}{dt}(\rho^2\dot{\varphi}) = 0.$$

The left-hand side in the second equation represents the acceleration of m in the direction perpendicular to the radius vector OP; it is equal to zero because there is no force in that direction. The equation immediately leads to the conclusion that $\rho^2\dot{\varphi} = \text{constant} = A$. If $\dot{\varphi}$ is determined from this equation and substituted in the first of the above equations, the final equation will contain only one dependent variable.

11. *Lagrange's equations*

In general the direct transformation of Newton's equations of motion from cartesian coordinates to the most convenient coordinates for the

solution of a given problem is laborious. For this reason, the transformation is usually carried out in general terms; the equations thus obtained,

$$\frac{d}{dt}\frac{\partial(T - V)}{\partial \dot{q}_m} - \frac{\partial(T - V)}{\partial q_m} = 0,$$

are called *Lagrange's equations*. In these equations, the q_m's are the *generalized coordinates* (R-2, p. 297 and R-3, p. 347); they represent a *minimum number* of independent variables needed to specify the configuration of the system. The functions T and V represent respectively the kinetic and potential energies of the system.

In polar coordinates, for example, the velocities of a particle in the ρ and φ directions are respectively $\dot{\rho}$ and $\rho\dot{\varphi}$; hence for a single particle

$$T = \tfrac{1}{2}m\dot{\rho}^2 + \tfrac{1}{2}m\rho^2\dot{\varphi}^2.$$

In the example of the preceding section the potential energy V at a point P is the work done *against* the force of gravity in bringing m to the point P from some fixed point at distance $\rho = a$ from M:

$$V = \int_a^\rho GMm\rho^{-2}\,d\rho = -GMm\rho^{-1} + GMma^{-1}.$$

Substituting these expressions in Lagrange's equations, we obtain the polar form for the equations of motion of m.

As another illustration let us consider a system consisting of two masses, M_1 and M_2, and two springs, S_1 and S_2, Figure 16.13. The kinetic and potential energies are

$$T = \tfrac{1}{2}M_1\dot{x}_1^2 + \tfrac{1}{2}M_2\dot{x}_2^2, \qquad V = \tfrac{1}{2}S_1x_1^2 + \tfrac{1}{2}S_2(x_2 - x_1)^2,$$

where x_1 and x_2 are the displacements from neutral positions. In the above expression for V it is assumed that no energy is stored in the springs when x_1 and x_2 are equal to zero. In the presence of gravity we must include other terms depending on the force of gravity; these terms, however, will only add constant terms to x_1 and x_2.

Substituting T, V in Lagrange's equations, we obtain the equations of motion,

$$M_1\ddot{x}_1 + (S_1 + S_2)x_1 - S_2x_2 = 0,$$
$$M_2\ddot{x}_2 - S_2x_1 + S_2x_2 = 0.$$

This particular set of equations can be obtained, almost as easily, directly from Newton's laws (R-6, 7). Lagrange's equations are a powerful tool in investigations of more complicated problems.

If there are externally applied forces F_m, not accounted for by the potential function, which tend to increase the corresponding generalized coordinates, then Lagrange's equations assume a more general form,

$$\frac{d}{dt}\frac{\partial(T-V)}{\partial \dot{q}_m} - \frac{\partial(T-V)}{\partial q_m} = F_m.$$

For other general principles of mechanics, such as Hamilton's Principle and the Principle of Least Action, consult R-2, 3, 8, 9.

12. *Lagrange-Maxwell equations*

Maxwell was the first to apply Lagrange's equations to electrical networks and thus obtain a set of circuit equations in terms of mesh currents. Since the electrical resistances are usually very important, another term must be included in Lagrange's equations to represent the effect of dissipation. Still another term is included to represent the externally applied electromotive forces; thus, Lagrange-Maxwell equations are expressed in the following form

$$\frac{d}{dt}\frac{\partial(T-V)}{\partial \dot{q}_m} - \frac{\partial(T-V)}{\partial q_m} + \frac{\partial \mathfrak{F}}{\partial \dot{q}_m} = E_m, \tag{3}$$

where q_m is the electric charge circulating in the mth mesh, $\dot{q}_m$ is the current circulating in the mth mesh, and E_m is the total external emf tending to increase $\dot{q}_m$. The functions T, V, $\mathfrak{F}$ are:

$$T = \tfrac{1}{2}\sum_{\alpha,}\sum_{\beta} L_{\alpha\beta}\dot{q}_\alpha\dot{q}_\beta, \qquad V = \tfrac{1}{2}\sum_{\alpha,}\sum_{\beta} \frac{q_\alpha q_\beta}{C_{\alpha\beta}}, \qquad \mathfrak{F} = \tfrac{1}{2}\sum_{\alpha,}\sum_{\beta} R_{\alpha\beta}\dot{q}_\alpha\dot{q}_\beta. \tag{4}$$

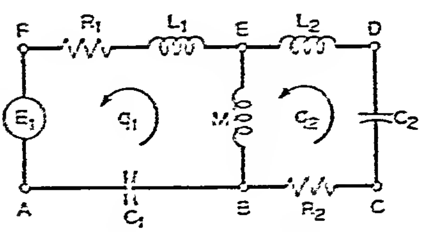

FIG. 22.3. A two-mesh electrical circuit.

If $\alpha \neq \beta$, the coefficients $L_{\alpha\beta}$, $C_{\alpha\beta}$, $R_{\alpha\beta}$ are called respectively the mutual inductance, the mutual capacitance, and the mutual resistance between the αth and βth meshes. If $\alpha = \beta$, the coefficients become the inductance, the capacitance, and the resistance of the corresponding mesh. For a more detailed discussion of these equations consult R-10.

To illustrate the application of the Lagrange-Maxwell equations, let us consider a two-mesh electrical circuit, Figure 22.3. As our generalized coordinates of the system we may choose the electric charges q_1 and q_2, circulating in the two meshes as indicated; the currents will be $\dot{q}_1$ and $\dot{q}_2$. It is obvious that we need at least two such coordinates; and it is also obvious that two are sufficient. The current from B to E is $\dot{q}_1 - \dot{q}_2$.

The magnetic energy T, the electric energy V, and the dissipation function $\mathfrak{F}$ are:

$$T = \tfrac{1}{2}L_1\dot{q}_1^2 + \tfrac{1}{2}M(\dot{q}_1 - \dot{q}_2)^2 + \tfrac{1}{2}L_2\dot{q}_2^2 = \tfrac{1}{2}(L_1 + M)\dot{q}_1^2 - M\dot{q}_1\dot{q}_2 + \tfrac{1}{2}(L_2 + M)\dot{q}_2^2$$

$$V = \frac{q_1^2}{2C_1} + \frac{q_2^2}{2C_2}, \qquad \mathfrak{F} = \tfrac{1}{2}R_1\dot{q}_1^2 + \tfrac{1}{2}R_2\dot{q}_2^2.$$

Substituting in (3), we have

$$(L_1 + M)\ddot{q}_1 - M\ddot{q}_2 + \frac{q_1}{C_1} + R_1\dot{q}_1 = E_1,$$

$$-M\ddot{q}_1 + (L_2 + M)\ddot{q}_2 + \frac{q_2}{C_2} + R_2\dot{q}_2 = 0.$$

13. *Kirchhoff's equations*

Another method of obtaining the equations for an electric network is to use Kirchhoff's laws (R-11, p. 173 and 460; R-4, pp. 101–102). The sum of the voltages round a closed circuit is zero, provided the impressed voltage is taken with the negative sign; hence in Figure 22.3,

$$V_{AB} + V_{BE} + V_{EF} - E_1 = 0, \qquad (5)$$
$$V_{BC} + V_{CD} + V_{DE} + V_{EB} = 0.$$

Let $\dot{q}_{FA}$ be the current flowing from F to A; the currents from A to B and from E to F are equal to $\dot{q}_{FA}$. Let $\dot{q}_{BC}$ be the current from B to C; the currents from C to D and from D to E are equal to $\dot{q}_{BC}$. Finally, let the current from B to E be $\dot{q}_{BE}$. Using the expressions for the voltages across the inductances and the resistances in terms of the currents, and for the voltages across the capacitances in terms of the charges, we may rewrite (5) as follows:

$$\frac{1}{C_1}\int^t \dot{q}_{FA}\,dt + M\ddot{q}_{BE} + L_1\ddot{q}_{FA} + R_1\dot{q}_{FA} - E_1 = 0, \qquad (6)$$

$$R_2\dot{q}_{BC} + \frac{1}{C_2}\int^t \dot{q}_{BC}\,dt + L_2\ddot{q}_{BC} - M\ddot{q}_{BE} = 0.$$

The sum of the currents entering (or leaving) each junction is zero; hence,

$$\dot{q}_{AB} + \dot{q}_{CB} + \dot{q}_{EB} = 0, \qquad \text{or} \qquad \dot{q}_{BE} = \dot{q}_{FA} - \dot{q}_{BC}. \qquad (7)$$

In their original form, Kirchhoff's equations contain more than the minimum number of coordinates needed to specify the state of the system. In the terminology of mechanics, the equations such as (7), arising from

the second Kirchhoff law, are always the *equations of constraint*, that is, simple relations between the coordinates. These equations are satisfied automatically if we express all branch currents in terms of mesh currents; then we obtain again the equations in the form given in the preceding section.

For further information on electric circuits the reader is referred to R-4, pp. 115–120; R-11, pp. 502–529; R-12.

14. *Compound interest and difference equations*

One of the simplest *difference equations* occurs in the calculation of compound interest and discount. Difference equations of the same type but of more complicated form are found in the theory of electric filters (R-4, 11, 13) and periodic structures in general (R-14).

Let y_n be the capital at the end of the nth interest period, *before* the interest has been added; then y_{n+1} is the capital at the beginning of the $(n+1)$th period, *after* the interest has been added, as well as the capital at the end of the $(n+1)$th period. The interest is proportional to the accumulated capital,

$$y_{n+1} - y_n = py_n.$$

From this equation we find that the ratio y_{n+1}/y_n is constant,

$$y_{n+1}/y_n = 1 + p;$$

therefore, starting with the initial value y_1, we obtain successive amounts

$$y_1, \qquad (1+p)y_1, \qquad (1+p)^2y_1, \qquad (1+p)^3y_1, \cdots.$$

In the case of compound discount p is negative.

15. *Periodic structures*

A simple periodic structure is illustrated by an infinite chain of equal mass particles M connected with springs whose stiffness is S, Figure 22.4.

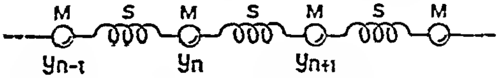

FIG. 22.4. A periodic structure consisting of mass particles and connecting springs.

Let y_n be the displacement of the nth particle to the right from its neutral position. The elastic restoring force to the right, tending to increase y_n, is $S(y_{n+1} - y_n)$; the elastic force to the left, tending to decrease y_n, is $S(y_n - y_{n-1})$; the difference, $S(y_{n+1} - 2y_n + y_{n-1})$, tends to increase y_n

and by Newton's second law

$$M\ddot{y}_n = S(y_{n+1} - 2y_n + y_{n-1}).$$

This is a *difference-differential* equation.

If we wish to investigate what happens when y_n is oscillating with the frequency ω radians per second, we let $\ddot{y}_n = -\omega^2 y_n$ since this is the general equation for such oscillations; thus we obtain a pure difference equation,

$$-\omega^2 M y_n = S(y_{n+1} - 2y_n + y_{n-1}). \tag{8}$$

In Section 14 we encountered a simpler difference equation, in which the ratio y_{n+1}/y_n was found to be constant. Let us see if the present equation possesses solutions of this type; that is, let us assume

$$y_{n+1}/y_n = y_n/y_{n-1} = k, \tag{9}$$

where k is constant. Dividing (8) by Sy_n, substituting from (9), and rearranging the terms, we have

$$k + (\omega^2 MS^{-1} - 2) + k^{-1} = 0, \qquad k^2 + (\omega^2 MS^{-1} - 2)k + 1 = 0. \tag{10}$$

Thus (8) possesses a solution in the form (9) provided k satisfies (10); that is,

$$k = (1 - \tfrac{1}{2}\omega^2 MS^{-1}) \pm \sqrt{(1 - \tfrac{1}{2}\omega^2 MS^{-1})^2 - 1}. \tag{11}$$

The product of these "characteristic values" of k is unity; so that if one value is k, the other is $1/k$. For each characteristic solution the displacement of the nth particle can thus be expressed in terms of the displacement of the zeroth particle:

$$y_n' = k^n y_0', \qquad y_n'' = k^{-n} y_0''. \tag{12}$$

By substitution it is easy to show that the sum of these characteristic solutions,

$$y_n = y_0' k^n + y_0'' k^{-n}, \tag{13}$$

is also a solution of (8). The constants y_0' and y_0'' are now the "arbitrary constants of summation."

Equation (8) may be considered as a recurrence formula for calculating any particular displacement from the two preceding displacements; therefore, if we assign arbitrary values to y_0 and y_1, we can obtain successively all the y's. Thus the most general solution contains two arbitrary constants, and equation (13) may be considered as one form of this general solution.

The constant k is real only if

$$|1 - \tfrac{1}{2}\omega^2 MS^{-1}| \geq 1, \qquad \tfrac{1}{2}\omega^2 MS^{-1} \geq 2, \qquad \omega \geq 2/\sqrt{MS^{-1}}. \tag{14}$$

If the magnitude of k is greater than unity, the magnitude of $1/k$ is automatically less than unity; the two characteristic solutions are essentially alike. One solution represents the increase in the amplitude of oscillations of the successive particles in the chain as we pass from left to right; and the other solution represents the same phenomenon as we pass from right to left.

If

$$\omega < 2\sqrt{MS^{-1}}, \tag{15}$$

k is complex; hence, y_n is complex. The actual displacements are real, of course. Since equation (8) is linear, the real part of y_n is also a solution. The convenience of complex variables for representing oscillations now becomes evident; then we can interpret the complex values of k directly. Thus if k is imaginary, equation (11) can be rewritten as

$$\begin{aligned} k &= (1 - \tfrac{1}{2}\omega^2 MS^{-1}) \pm i\sqrt{1 - (1 - \tfrac{1}{2}\omega^2 MS^{-1})^2} \\ &= \cos\vartheta \pm i \sin\vartheta, \end{aligned} \tag{16}$$

where

$$\begin{aligned} \cos\vartheta &= 1 - \tfrac{1}{2}\omega^2 MS^{-1}, \qquad 1 - \cos\vartheta = \tfrac{1}{2}\omega^2 MS^{-1}, \\ \sin(\vartheta/2) &= \tfrac{1}{2}\omega\sqrt{MS^{-1}}. \end{aligned} \tag{17}$$

The two characteristic values are now k and its conjugate k^*; since their product $kk^* = 1$, their absolute values are equal to unity. Hence, multiplication by k affects only the phase of oscillations; for frequencies satisfying the inequality (15), the oscillations are propagated along the chain without "attenuation."

By referring to R-14, the reader will catch a glimpse of the extremely wide range of application of the ideas introduced in this section.

16. *Electrical filters*

A chain of coupled electric circuits, Figure 22.5, represents another example of a periodic structure. Assuming steady state oscillations and applying the first Kirchhoff law to a typical mesh, we have

$$-Z_M I_{n-1} + Z I_n - Z_M I_{n+1} = 0,$$

where Z_M is the mutual impedance between the adjacent meshes, Z the self-impedance of a mesh, and I_n the current in the nth mesh. This difference equation is similar to the one in the preceding section. Rewriting

it in a symmetrical form we have

$$\frac{I_{n+1}}{I_n} + \frac{I_{n-1}}{I_n} = \frac{Z}{Z_M}.$$

The current ratio is denoted by k and its permissible values are obtained from a quadratic equation.

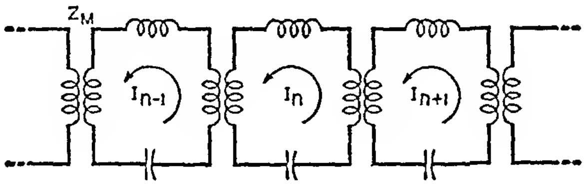

FIG. 22.5. A periodic structure consisting of a chain of coupled electric circuits.

In the above example we assumed that only the adjacent meshes were coupled. When the interaction is extended to the second neighbors, then

$$-Z''_M I_{n-2} - Z'_M I_{n-1} + Z I_n - Z'_M I_{n+1} - Z''_M I_{n+2} = 0.$$

There will be no change in the method of solution; only the characteristic equation will be of higher degree.

17. *Bending of beams*

If the cross section of a beam is small compared with the length, the problem of obtaining the shape of the beam under a given load consists largely in the derivation of the appropriate equations. The actual solution of the equations is extremely simple, at least in many important cases. Thus in R-7, p. 221; R-15, p. 135; R-16, p. 1; or R-17, pp. 268–269, the differential equation is given as

$$EI\frac{d^2y}{dx^2} = -M, \tag{18}$$

where E is Young's modulus, I is the moment of inertia, M is the bending moment, and y is the deflection of a point whose coordinate is x. The shearing stress S is given by

$$S = \frac{dM}{dx} = -\frac{d}{dx}\left(EI\frac{d^2y}{dx^2}\right), \tag{19}$$

and the load is equal to

$$p(x) = -\frac{dS}{dx} = \frac{d^2}{dx^2}\left(EI\frac{d^2y}{dx^2}\right). \tag{20}$$

Hence if the beam is uniform and its weight per unit length is W, the differential equation for the elastic curve becomes

$$\frac{d^2}{dx^2}\left(EI\frac{d^2y}{dx^2}\right) = -W. \tag{21}$$

This equation may be solved by successive integrations. Integrating once, we have

$$\frac{d}{dx}\left(EI\frac{d^2y}{dx^2}\right) = -Wx + C_1, \tag{22}$$

where C_1 is a constant of integration which must be determined from the boundary conditions. If one end of the beam, $x = 0$, is clamped and the other, $x = l$, is free,

$$y(0) = 0, \qquad y'(0) = 0, \qquad y''(l) = 0, \qquad y'''(l) = 0. \tag{23}$$

The first equation expresses the condition that the displacement of the clamped end must vanish, the second that the elastic curve there must be tangent to its undistorted line, the third equation states that the curvature of the elastic curve must be zero at the free end, and the fourth that the shearing force must also vanish at the free end. Using the last of these equations, we find $C_1 = Wl$. Substituting in (22) and integrating once more, we have

$$EI\frac{d^2y}{dx^2} = -\tfrac{1}{2}W(x-l)^2 + C_2.$$

To satisfy the next to the last of the set of boundary conditions, we must take $C_2 = 0$. Then,

$$EI\frac{dy}{dx} = -\frac{1}{6}W(x-l)^3 + C_3, \qquad \frac{1}{6}Wl^3 + C_3 = 0;$$

$$EIy = -\tfrac{1}{24}W(x-l)^4 - \tfrac{1}{6}Wl^3x + C_4,$$

$$0 = -\tfrac{1}{24}Wl^4 + C_4;$$

and finally,

$$\begin{aligned} EIy &= -\tfrac{1}{24}W(x-l)^4 - \tfrac{1}{6}Wl^3x + \tfrac{1}{24}Wl^4 \\ &= -\tfrac{1}{24}Wx^4 + \tfrac{1}{6}Wlx^3 - \tfrac{1}{4}Wl^2x^2. \end{aligned}$$

If there is an additional load W_1 distributed between $x = \xi - \frac{1}{2}s$ and $x = \xi + \frac{1}{2}s$, then (21) is true everywhere outside the interval $(\xi - \frac{1}{2}s, \xi + \frac{1}{2}s)$; but inside the interval we have

$$\frac{d^2}{dx^2}\left(EI\frac{d^2y}{dx^2}\right) = -W - W_1.$$

Integrating over this interval,

$$\frac{d}{dx}\left(EI\frac{d^2y}{dx^2}\right)\Big|_{\xi-\frac{1}{2}s}^{\xi+\frac{1}{2}s} = -\int_{\xi-\frac{1}{2}s}^{\xi+\frac{1}{2}s} W\,dx - \int_{\xi-\frac{1}{2}s}^{\xi+\frac{1}{2}s} W_1\,dx. \tag{24}$$

The last integral represents the total additional weight P. If P is highly concentrated, the first integral on the right may be neglected and (24) may be replaced by the following discontinuity condition

$$\frac{d}{dx}\left(EI\frac{d^2y}{dx^2}\right)\Big|_{\xi-0}^{\xi+0} = -P. \tag{25}$$

Boundary conditions may be varied as illustrated by the sketches on pp. 60–61 of R-18. See also pp. 63–66 of R-19. Both books are excellent introductions to elementary differential equations.

18. *Vibrations of strings*

Consider a string under tension T, Figure 22.6. The " tension " is the force stretching the string in its neutral position AB. Confining ourselves to " small amplitude " oscillations, we shall neglect the terms depending

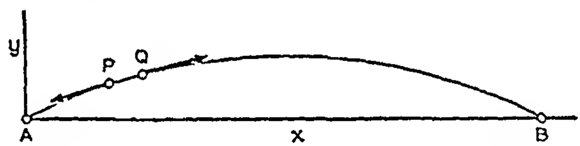

FIG. 22.6. Forces acting on an element of a string under tension.

on the square and the higher powers of the displacement y of a differential element PQ of the string. In the displaced position the string becomes longer and the tension must increase; however, the increase in length depends on the square of y_{max} and we have decided to neglect quantities of this order of magnitude. If α is the inclination of PQ to the horizontal, the horizontal component of the tension is $T\cos\alpha \simeq T(1-\frac{1}{2}\alpha^2)$. Again we neglect the small quantity α^2; this means that we ignore the horizontal movement of PQ and concern ourselves with the vertical movement only. The vertical component of tension is $T\sin\alpha$; but for small α, $\sin\alpha$ is approximately equal to $\tan\alpha$ and $\tan\alpha = \partial y/\partial x$. Thus the vertical component of the tension at PQ is $T\,\partial y/\partial x$ and the *vertical force* tending to increase y is $\Delta(T\,\partial y/\partial x)$. By Newton's second law this must equal the mass $M\Delta x$ of PQ times the acceleration $\partial^2 y/\partial x^2$. Hence, letting Δx approach zero, we have

$$\frac{\partial}{\partial x}\left(T\frac{\partial y}{\partial x}\right) = M\frac{\partial^2 y}{\partial x^2},$$

where M is the mass per unit length.

19. *Poisson's equation*

The theory of static electric fields is based on two experimental equations:

$$\oint E_s\,ds = 0, \qquad \int\int \epsilon E_n\,dS = Q. \tag{26}$$

The first equation expresses the fact that the line integral of the tangential component of the electric intensity (defined as the force per unit charge) round a closed curve vanishes; the second equation states that the surface integral of the normal component of E taken over a closed surface is proportional to the enclosed charge. The coefficient of proportionality ϵ depends on the medium. These equations were suggested by experimental facts; but in view of the necessarily limited number of experiments they are essentially postulates.

From the definitions of the curl and the divergence of a vector (see Chapter 7) and from (26) we have

$$\text{curl } E = 0, \qquad \text{div } (\epsilon E) = q, \tag{27}$$

where q is the volume charge density. The first of these equations yields

$$E = -\text{ grad } V, \tag{28}$$

where V is a scalar function. Substituting in the second equation, we have

$$\text{div } (\epsilon \text{ grad } V) = -q. \tag{29}$$

Expanding,

$$\epsilon \Delta V + \text{grad } \epsilon \cdot \text{grad } V = -q. \tag{30}$$

If the medium is homogeneous, ϵ is constant and

$$\Delta V = -q/\epsilon. \tag{31}$$

This is *Poisson's equation.*

If $q = 0$, we obtain *Laplace's equation*

$$\Delta V = 0. \tag{32}$$

These equations are very general. Thus the first equation in the set (26) applies to any *conservative field of force*, that is, to a field of force in which the net work done on a particle moving completely round a closed curve vanishes. The second equation in the set (26) is the "equation of continuity." The vector ϵE may represent, for instance, the rate of flow per unit area and Q will then be the rate of decrease of the substance in the enclosed volume.

20. *Maxwell's equations*

In the case of variable electromagnetic fields the postulates suggested by experimental facts are:

$$\oint E_s\, ds = -\frac{\partial}{\partial t}\int\int \mu H_n\, dS,$$
$$\oint H_s\, ds = \int\int gE_n\, dS + \frac{\partial}{\partial t}\int\int \epsilon E_n\, dS, \tag{33}$$

where E and H are the electric intensity and the magnetic intensity at a point. The parameters μ, g, ϵ define the properties of the medium.

Using the definition of the curl of a vector, we obtain

$$\operatorname{curl} E = -\frac{\partial(\mu H)}{\partial t}, \qquad \operatorname{curl} H = gE + \frac{\partial(\epsilon E)}{\partial t}. \tag{34}$$

If we apply (33) directly to differential curvilinear rectangles formed by coordinate lines, we shall obtain the differential equations in the various coordinate systems.

21. *On applications of differential equations*

Aside from elementary mathematics, differential equations enjoy the widest field of applications. Derivatives are rates of change and as such they enter our thinking about all kinds of natural phenomena. Certain differential equations occur in so many different fields that they are labeled with distinctive names, such as the wave equation, Laplace's equation, Bessel's equation, etc. In Section 18 we derived the one-dimensional wave equation for vibrating strings. The same equation applies to longitudinal vibrations of air in tubes, to electric waves on wires, to "waves of probability" in wave mechanics; only the physical meaning of the variables is different in each case.

Most differential equations are not solvable in "closed form," that is, in a form containing a finite number of "known" functions; but some are solvable in this sense and more become solvable as the number of "known" functions is extended. It is always advisable to try different transformations of the dependent and independent variables that may reduce a given equation to some other equation whose solution is known. For these "elementary methods" the reader is referred to R-18 and R-19. For the more advanced theory we suggest R-20, 21, 22, 23. Perturbation methods in the form particularly suitable to the analysis of *free* oscillations of nonuniform strings and to problems of atomic physics are explained in

R-6, 24, 25. Some of these problems, however, may be treated more successfully by the " wave perturbation method " explained in this book. In connection with partial differential equations we suggest R-26, 27. For numerical methods of solving ordinary differential equations, see R-28, and for partial differential equations R-29.

Then there are more specialized treatises dealing with one special type of differential equation. Thus, R-30 is a treatise on differential equations with constant coefficients applied to vibrating systems, mechanical and electrical. Mathematical unification of electrical and mechanical systems is particularly well exhibited in R-31.

For the applications of the more general types of differential equations, see R-32, 33, 34, 36.

22. *On applications of complex variables*

Next to differential equations, the theory of analytic functions (or functions of a complex variable) has perhaps the widest range of application. First of all there are the exceedingly useful applications to the theory of steady state electrical and mechanical oscillations. Electrical engineers have used this method most; but it is spreading to other fields. In its elementary aspects (see Section 1.10), the method is essentially a shorthand representation of sinusoidal functions of the same frequency. Such functions are completely determined by the amplitude and the initial phase and thus can be represented either by two-dimensional vectors or by complex numbers. The amplitude-phase relationships can thus be conveniently exhibited. The addition of sinusoidal functions is reduced to the addition of vectors or to the addition of complex numbers. The relationship between two sinusoids can be expressed as the ratio of the complex numbers representing these sinusoids, that is, as a complex number. These considerations lead to the idea of " complex impedance " as a generalized resistance. Corresponding to Ohm's " law " of proportionality between a constant voltage and current and Hooke's " law " of proportionality between a small constant cause and its effect we shall have the generalized Ohm's law and Hooke's law for steady state oscillations.

While the original objective in applying complex numbers to steady state oscillations was modest, it soon became apparent that the theory of linear electric circuits is essentially a branch of the theory of functions of the complex variable (see R-37). The same is true of linear electromechanical systems (R-38). Nyquist's criterion of stability of dynamical systems is expressed as a condition on the location of the zeros and poles of a function in the complex plane (R-38).

Another series of exceedingly important applications is based on the relationship between functions of a complex variable and the two-dimensional Laplace's equation

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} = 0. \tag{35}$$

Introducing new independent variables

$$u = x + iy, \qquad v = x - iy, \tag{36}$$

in (35), we obtain

$$\frac{\partial^2 W}{\partial u \partial v} = 0. \tag{37}$$

Thus W can be either an arbitrary function of u (since the partial derivative of such a function with respect to v vanishes) or an arbitrary function of v or the sum of such functions,

$$W = F(u) + G(v) = F(x + iy) + G(x - iy). \tag{38}$$

From this result stem the applications to electrostatics and magnetostatics (R-4, 10), to hydrodynamics (R-39, 40, 41), to aerodynamics (R-42, 43). The famous Joukowski's profiles of wings are obtained by means of certain transformations of functions of a complex variable.

The representation of the unit step function and the unit impulse function by certain integrals in the complex plane, the Laplace integrals, opens a way to the applications of functions of a complex variable to transient oscillations (Chapter 16 of this book and R-30). Laplace integrals are also useful in the theory of waves (R-4, 44).

23. *Matrix algebra*

Systematization of rules of manipulation is an essential function of mathematics. Suppose for instance that we have a chain of "electrical transducers" or "four-terminal networks," Figure 22.7. In each box we have a linear electric network. For each of these networks, no matter how complicated, the complex voltage V and current I at one pair of terminals are linear functions of the voltage and current at the other pair of terminals,

$$\begin{aligned} V_2 &= aV_1 + bI_1, \qquad & V_3 &= mV_2 + nI_2, \\ I_2 &= cV_1 + dI_1, \qquad & I_3 &= pV_2 + qI_2. \end{aligned} \tag{39}$$

To express V_3, I_3 in terms of V_1, I_1 we need only substitute the first column

of the above equations in the second,

$$\begin{aligned} V_3 &= m(aV_1 + bI_1) + n(cV_1 + dI_1), \\ I_3 &= p(aV_1 + bI_1) + q(cV_1 + dI_1). \end{aligned} \tag{40}$$

Collecting the terms,

$$\begin{aligned} V_3 &= (ma + nc)V_1 + (mb + nd)I_1, \\ I_3 &= (pa + qc)V_1 + (pb + qd)I_1. \end{aligned} \tag{41}$$

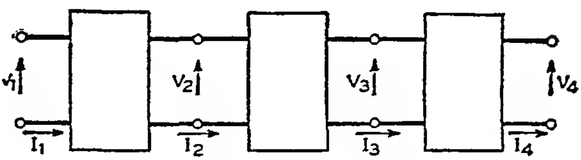

FIG. 22.7. A chain of four-terminal transducers.

If we had to perform such substitutions only occasionally, we would feel no need either for any labor saving scheme or for a shorthand notation; but if we have to do it often, and many times in the same problem, then we shall probably want to look for rules which will enable us to omit some of the steps in the manipulation. If we examine the coefficients in equations (41), we find that they are the sums of the products of the coefficients in the rows of the second group of equations (39) and the corresponding coefficients in the columns of the first group. This suggests that the intermediate step (40) may be omitted if we form a rule for obtaining the coefficients of (41) from those of (39). In fact, the formation of the pair of quantities (V_2,I_2) from the pair (V_1,I_1) promises to follow the same rule if we write

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}. \tag{42}$$

Thus V_2 is obtained if we multiply the first row of the bracketed set of coefficients on the right by the first (and in this case the only) column of the second bracketed set. Similarly, I_2 is obtained by multiplying the second row by the first column. Let us now write the second pair of equations in (39) as

$$\begin{bmatrix} V_3 \\ I_3 \end{bmatrix} = \begin{bmatrix} m & n \\ p & q \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}. \tag{43}$$

Suppose we substitute from (42) in (43), treating the bracketed sets of

quantities as if they were ordinary algebraic quantities:

$$\begin{bmatrix} V_3 \\ I_3 \end{bmatrix} = \begin{bmatrix} m & n \\ p & q \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}. \tag{44}$$

It is tedious to have to repeat such a long phrase as "bracketed set of quantities"; it is better to give it a short name "matrix." The matrix equation (44) will be equivalent to (41) provided we define the "product of two matrices" as follows:

$$\begin{bmatrix} m & n \\ p & q \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ma + nc & mb + nd \\ pa + qc & pb + qd \end{bmatrix}. \tag{45}$$

The rule is to multiply the elements in the rows of the first matrix by the elements in the columns of the second and add the results. The position of each sum in the product matrix is given by the serial numbers of the row and the column which were multiplied together.

More generally the product of a matrix with n columns and a matrix with n rows is defined as follows:

$$\begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \cdots & b_{1k} \\ b_{21} & b_{22} \cdots & b_{2k} \\ \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} \cdots & b_{nk} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \cdots & c_{1k} \\ c_{21} & c_{22} \cdots & c_{2k} \\ \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} \cdots & c_{mk} \end{bmatrix}, \tag{46}$$

where

$$c_{ij} = \sum_{\alpha} a_{i\alpha} b_{\alpha j} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots a_{in}b_{nj}.$$

Thus the elements in the ith row of the first matrix are multiplied by the corresponding elements in the jth column of the second matrix; then these products are added and the result is written as the element in the ith row and jth column of the product matrix. This rule of multiplication requires that the number of elements in a row of the first factor be equal to the number of elements in a column of the second factor.

It is easy to see that in general the matrix product is not commutative,

$$AB \neq BA. \tag{47}$$

For more information about matrices, see R-45.

24. *Functions in the making*

When confronted with a new function, the student is apt to feel a little mystified. The simple fact is that when a certain function occurs frequently, it is convenient to tabulate it and then publish the tables for general use. This requires a name and a symbol. Suppose we read:

"The *integral sine*, Si x, and the *integral cosine*, Ci x, are defined as follows:

$$\text{Si}\, x = \int_0^x \frac{\sin t}{t}\, dt, \qquad \text{Ci}\, x = \int_\infty^x \frac{\cos t}{t}\, dt.\text{"} \tag{48}$$

The definition is complete and, as a rule, we should not expect much more than a table. Some functions will possess many remarkable properties; but this is not true of all functions. The functions defined by (48) are just ordinary integrals; these integrals are given special names simply because they occur frequently in important problems.

Si x and Ci x occur in the theory of radiation, in diffraction theory, in antenna theory, and in the solution of linear differential equations by the wave perturbation method. In the latter case certain integrals involving Si x and Ci x become important. Some of these integrals can be evaluated in "closed form" provided we regard Si x and Ci x as "known functions"; but other integrals represent "new functions." Recently the SC-function has been defined by Miss Marion C. Gray as follows:

$$\text{SC}(x) = \int_0^x \text{Si}\, u\, d\text{Ci}\, u = \int_0^x \text{Si}\, u \frac{\cos u}{u}\, du.$$

Numerical values of this integral are needed in recent developments in antenna theory, and may be obtained from the expansion

$$\text{SC}(x) = \sum_{n=0}^{\infty} \frac{(-)^n x^{2n+1}}{(2n+1)!}\left(1 + \frac{1}{3^2} + \cdots + \frac{1}{(2n+1)^2}\right)\left(\cos x + \frac{x \sin x}{2n+2}\right).$$

For large values of x we can also find an asymptotic expansion

$$\begin{aligned}\text{SC}(x) \sim \text{Si}\, x\, \text{Ci}\, x &+ \frac{1}{2x}\left(1 - \frac{2!}{3x^2} + \frac{4!}{5x^4} - \cdots\right) \\ &+ \frac{\sin 2x}{4x^2}\left(1 - \frac{5}{x^2} + \frac{64}{x^4} - \cdots\right) \\ &- \frac{\cos 2x}{2x^3}\left(1 - \frac{8}{x^2} + \frac{156}{x^4} - \cdots\right).\end{aligned}$$

A short table of numerical values of SC($k\pi$) may be of interest:

k	SC($k\pi$)	k	SC($k\pi$)
0.2	0.58375	1.2	−0.03610
.4	.93065	1.4	− .19388
.6	.93372	1.6	− .19628
.8	.65818	1.8	− .09587
1.0	.27784	2.0	.04469

25. Dimensional analysis

Any physical equation must be "dimensionally correct." An equation such as $\ell = t$, where ℓ represents length and t is time, is meaningless. If the side of a square is 3 meters and a certain interval of time is 3 seconds, the numbers expressing the size of the two physical quantities happen to be equal only as long as we measure one quantity in meters and the other in seconds. As soon as we change the units, the numbers cease to be equal. Equations relating physical quantities must be true regardless of the choice of units. For this reason, in solving a physical problem, it is advisable to check from time to time the dimensions of the various terms in the equations; some errors may thus be caught in the early stages of the solution.

Dimensional considerations may be used to obtain information about functional relationships between the various physical quantities in a given problem. Let us consider the classical example of a pendulum. Assume that the length of the rod (weightless) is ℓ, the mass on the end is m, and the period of oscillations is t. It is easily seen that a relation of the form

$$t = k\ell^{\alpha}m^{\beta}, \tag{49}$$

where k is a dimensionless constant and α, β are unknown exponents, cannot possibly be true, since length, mass, and time are dimensionally independent. Then it may occur to us that the oscillations are sustained by the force of gravity, mg, where g is the acceleration of gravity; so that perhaps we should assume that t depends on mg and not on m. Thus let us write

$$t = k\ell^{\alpha}(mg)^{\beta}. \tag{50}$$

The dimensional formula for g is

$$[g] = [\ell t^{-2}], \tag{51}$$

where the brackets are used to remind us that in (51) we are thinking only about the qualitative variation of the acceleration with length and time. The dimensional equation for (50) is then

$$[t] = [\ell^{\alpha}m^{\beta}\ell^{\beta}t^{-2\beta}]. \tag{52}$$

This requires

$$1 = -2\beta, \qquad \alpha + \beta = 0, \qquad \beta = 0. \tag{53}$$

These equations have no solution. On further consideration it may occur to us that the period of oscillation may depend in one way on the mass of the pendulum and in another way on the force of gravity; or in other words, the period may contain m and g independently. Thus let us write

$$t = k\ell^{\alpha}m^{\beta}g^{\gamma}. \tag{54}$$

The dimensional equation is then

$$[t] = [\ell^{\alpha} m^{\beta} \ell^{\gamma} t^{-2\gamma}] = [\ell^{\alpha+\gamma} m^{\beta} t^{-2\gamma}]. \tag{55}$$

This requires

$$1 = -2\gamma, \qquad \alpha + \gamma = 0, \qquad \beta = 0. \tag{56}$$

Therefore,

$$\gamma = -\tfrac{1}{2}, \qquad \alpha = -\gamma = \tfrac{1}{2}, \qquad \beta = 0. \tag{57}$$

Hence, (54) becomes

$$t = k\sqrt{\ell/g}, \tag{58}$$

where k is a dimensionless constant of proportionality. This is the true equation; for oscillations of small amplitude we already know that $k = 2\pi$. The dimensional analysis can give us no information about the functional dependence of t on the angular amplitude of oscillations since angles are dimensionless.

Next, let us analyze an electric circuit consisting of an inductance L and capacitance C in series. Assuming that the period of oscillations depends on L and C, we write

$$t = kL^{\alpha}C^{\beta}. \tag{59}$$

If we regard the electric charge, q, as having a physical dimension of its own, independent of mass, length, and time, then

$$\begin{aligned}[C] &= [q^2 m^{-1} \ell^{-2} t^2], \\ [L] &= [q^{-2} m \ell^2].\end{aligned} \tag{60}$$

The dimensional equation for (59) is then

$$[t] = [q^{-2\alpha+2\beta} m^{\alpha-\beta} \ell^{2\alpha-2\beta} t^{2\alpha}]. \tag{61}$$

Hence,

$$\alpha = \tfrac{1}{2}, \qquad \beta = \tfrac{1}{2}, \tag{62}$$

and (59) becomes

$$t = k\sqrt{LC}. \tag{63}$$

Let us now include a resistance R in the circuit. Dimensional relationships between R, L and C may be expressed as follows:

$$[L] = [Rt], \qquad [C] = [R^{-1}t]. \tag{64}$$

Assuming

$$t = kL^{\alpha}C^{\beta}R^{\gamma} \tag{65}$$

for the period of oscillations, we have

$$[t] = [R^{\alpha-\beta+\gamma}t^{\alpha+\beta}],$$

$$\alpha - \beta + \gamma = 0, \qquad \alpha + \beta = 1, \tag{66}$$

$$\alpha = \tfrac{1}{2} - \tfrac{1}{2}\gamma, \qquad \beta = \tfrac{1}{2} + \tfrac{1}{2}\gamma.$$

Therefore

$$t = k\sqrt{LC}\,(R\sqrt{C/L})^{\gamma}. \tag{67}$$

The exponent γ is left arbitrary. Furthermore, the quantity in parentheses is dimensionless so that we could add arbitrary powers together and form an arbitrary function. Equation (67), therefore, tells us no more than

$$t = \sqrt{LC}\,f(R\sqrt{C/L}), \tag{68}$$

where f is an arbitrary function. This time we do not obtain as much information as in the preceding example. This is not surprising since we can insert R either in series with L and C, or in parallel with L, or in parallel with C; and in each case the result will be different. But we shall always find that R is associated with the same factor $\sqrt{C/L}$.

Suppose that the circuit contains two inductances L_1, L_2 and a capacitance C; then

$$t = kL_1^{\alpha}L_2^{\beta}C^{\gamma}. \tag{69}$$

In this case,

$$\alpha + \beta + \gamma = 1, \qquad \alpha + \beta - \gamma = 0;$$

$$\gamma = \tfrac{1}{2}, \qquad \alpha + \beta = \tfrac{1}{2}, \qquad \alpha = \tfrac{1}{2} - \beta.$$

Therefore,

$$t = k\sqrt{L_1C}(L_2/L_1)^{\beta},$$

with an arbitrary β; hence,

$$t = \sqrt{L_1C}\,f(L_2/L_1). \tag{70}$$

It is evident that the dimensional analysis is ambiguous. In the case of the pendulum problem we got into difficulties at first because we failed to include a sufficient number of independent quantities on which the period might depend. If we include more quantities than are needed, the answer will become less informative. If, for instance, we assume that the period of oscillation of the electric circuit depends not only on the inductance L and the capacitance C but also on the amplitudes, q and I, of the electric charge on the capacitor and the electric current through the inductor, then

instead of (63) we shall obtain

$$t = \sqrt{LC} f(I\sqrt{LC}/q), \tag{71}$$

where f is an arbitrary function. Unless we know that t does not depend on q, we cannot really neglect the possibility that it might. If we make a similar supposition in the case of the pendulum, we shall have to conclude that the dimensionless factor k in (58) is a constant; but actually it is not. There we decided that k could depend on the maximum angular displacement because the latter is dimensionless. A more straightforward reason would be: equation (54) may possibly contain a factor s^δ where s is the maximum linear displacement of the mass. Then instead of (58) we obtain

$$t = \sqrt{\ell/g}\, f(s/\ell). \tag{72}$$

The argument of f is the angular amplitude.

In gaussian units the dimensions of the inductance and capacitance are those of length. In these units (59) is an impossible equation. What other quantity should we include? It is doubtful that we could think of the proper quantity if we did not already know that in gaussian units the velocity of light v is apt to appear in all sorts of unexpected places. Including the velocity of light, we would obtain

$$t = \frac{L}{v} f(C/L) = \frac{C}{v} f_1(L/C) = \frac{\sqrt{LC}}{v} f_2(L/C). \tag{73}$$

These are less informative equations than (63).

Dimensional considerations are useful in devising model experiments. Suppose we are unable to obtain the formula for the impedance of a broadcast antenna from theory and have to depend on experimental measurements. Suppose we want this information for design purposes. It would cost a great deal to build a large number of towers of different dimensions in order to obtain enough information for our purposes. If we could obtain the needed information from measurements on small scale models, it would not cost nearly so much. In order to calculate the impedance we have to solve Maxwell's equations:

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -i\omega\mu H_z, \qquad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = i\omega\epsilon E_z,$$

and four additional equations which may be obtained by the cyclic permutation of x, y, z. Dividing by ω, we obtain

$$\frac{\partial E_y}{\partial(\omega x)} - \frac{\partial E_x}{\partial(\omega y)} = -i\mu H_z, \qquad \frac{\partial H_y}{\partial(\omega x)} - \frac{\partial H_x}{\partial(\omega y)} = i\epsilon E_z. \tag{74}$$

If we have reason to believe that the resistance of the antenna and the ground is negligible for our purposes, then we have to solve equations (74) in a single homogeneous medium where μ and ϵ are constants. The solution must also be such that the component of E tangential to the surface of the antenna vanishes. Under these conditions our solution will be a function of ωx, ωy, ωz, μ, ϵ and will remain the same as long as the products of the frequency $f = \omega/2\pi$ and the dimensions of the antenna remain the same. That is, we can scale the antenna down to small dimensions and compensate for this by raising the frequency in the same ratio.

For more information about dimensional analysis, see R-46.

26. *Concluding remarks on applicable mathematics*

Applications of mathematics to science depend on the similarity between physical and mathematical processes. This analogy or isomorphism enables us to substitute in our thinking a mathematical process for a physical. Conversely, every physical process " solves " a mathematical problem. It is a " lend-lease " that works in both directions: on the one side we have applied mathematics and on the other analogue calculating machines. Diverse physical processes may be mathematically alike if we ignore their special physical characteristics. Oscillations and waves in many mechanical and electrical systems obey the same equations. This enables us to formulate a general theory of oscillations and waves and thus save time and effort; it also makes possible mechanical solutions of electrical problems or electrical solutions of mechanical problems. We can make analogue experiments as well as analogue calculating machines.

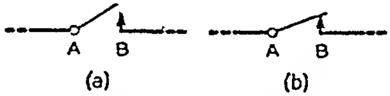

FIG. 22.8. (a) An open relay; (b) A closed relay.

What kind of mathematics is applicable? All mathematics is probably applicable, although some more frequently applied than others. An interesting example of an unusual application may be found in R-47 where George Boole developed the "algebra of logic." The same algebra is also applicable to electrical relay circuits. A simple relay is just a switch which can be either closed or open. If the switch is open as in Figure 22.8(a), the current is not permitted to flow from A to B; otherwise, it is. Imagine now a number system consisting of only two numbers, 0 and 1. Let the rules for the addition and multiplication of these numbers be:

$$0 + 0 = 0, \qquad 1 + 0 = 0 + 1 = 1, \qquad 1 + 1 = 1;$$

$$0 \cdot 0 = 0, \qquad 1 \cdot 0 = 0 \cdot 1 = 0, \qquad 1 \cdot 1 = 1.$$

Let the " admittance " x of a relay be 1 when it is closed and 0 when it is open. Consider two relays whose admittances are x and y. The admittance z of the series connection of these relays, Figure 22.9(a), is then the product of their admittances,

$$z = xy,$$

for it requires the closure of both to complete the circuit. The admittance z of two relays connected in parallel as in Figure 22.9(b) is the sum of their admittances,

$$z = x \dotplus y,$$

for the closure of either will complete the circuit.

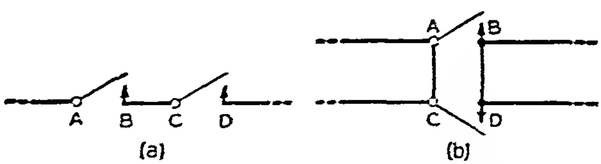

FIG. 22.9. (a) Two relays *in series*; (b) Two relays *in parallel.*

We could have defined the " impedance " of a relay in such a way that it is unity when the relay is open, and zero when the relay is closed; then we should find that the impedance of the relays in series is the sum of their impedances, and the impedance of a parallel combination is the product of their impedances.

The application of this peculiar arithmetic to relay circuits first occurred to Claude E. Shannon. For more detail see his paper on " A Symbolic Analysis of Relay and Switching Circuits," A.I.E.E., *Trans.*, Vol. 57, 1938.

REFERENCES

1. Leigh Page, *Introduction to Theoretical Physics*, D. Van Nostrand Company, Inc., New York, Second Edition, 1935.

2. William Fogg Osgood, *Mechanics*, The Macmillan Company, New York, 1937.

3. William Duncan MacMillan, *Theoretical Mechanics — Statics and the Dynamics of a Particle*, McGraw-Hill Book Company, Inc., New York, 1927.

4. S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand Company, Inc., New York, 1943.

5. Frederick H. Getman, *Outlines of Theoretical Chemistry*, John Wiley and Sons, Inc., New York, Fourth Edition, 1928.

6. Philip M. Morse, *Vibration and Sound*, McGraw-Hill Book Company, Inc., New York, 1936.

7. S. Timoshenko, *Vibration Problems in Engineering*, D. Van Nostrand Company, Inc., New York, 1928.

8. Ernst Mach, *The Science of Mechanics*, The Open Court Publishing Company, Inc., Chicago and London, 1919.

9. William Elwood Byerly, *Introduction to the Calculus of Variations*, Harvard University Press, Cambridge, 1928.

10. James Jeans, *The Mathematical Theory of Electricity and Magnetism*, Cambridge University Press, Fifth Edition, 1933.

11. Leigh Page and Norman Ilsley Adams, *Principles of Electricity*, D. Van Nostrand Company, Inc., New York, 1931.

12. Ernst A. Guillemin, *Communication Networks*, Vol. I, *The Classical Theory of Lumped Constant Networks*, John Wiley and Sons, Inc., New York, 1931.

13. Ernst A. Guillemin, *Communication Networks*, Vol. II, *The Classical Theory of Long Lines, Filters and Related Networks*, John Wiley and Sons, Inc., New York, 1935.

14. Leon Brillouin, *Wave Propagation in Periodic Structures*, McGraw-Hill Book Company, Inc., New York, 1946.

15. S. Timoshenko, *Strength of Materials*, Part I, *Elementary Theory and Problems*, D. Van Nostrand Company, Inc., New York, Second Edition, 1940.

16. S. Timoshenko, *Strength of Materials*, Part II, *Advanced Theory and Problems*, D. Van Nostrand Company, Inc., New York, Second Edition, 1941.

17. Theodore Von Karman and Maurice Biot, *Mathematical Methods in Engineering*, McGraw-Hill Book Company, Inc., New York, 1940.

18. Donald Francis Campbell, *A Short Course on Differential Equations*, The Macmillan Company, New York, 1926.

19. H. B. Phillips, *Differential Equations*, John Wiley and Sons, Inc., New York, 1924.

20. H. Bateman, *Differential Equations*, Longmans, Green and Co., Ltd., London, 1926.

21. H. T. H. Piaggio, *An Elementary Treatise on Differential Equations and their Applications*, G. Bell and Sons, Ltd., London, 1920.

22. F. R. Moulton, *Differential Equations*, The Macmillan Company, New York, 1930.

23. E. L. Ince, *Ordinary Differential Equations*, Longmans, Green and Co., Ltd., London, 1927.

24. John C. Slater and Nathaniel H. Frank, *Introduction to Theoretical Physics*, McGraw-Hill Book Co., Inc., New York, 1933.

25. Henry Margenau and George Moseley Murphy, *The Mathematics of Physics and Chemistry*, D. Van Nostrand Company, Inc., New York, 1943.

26. Arthur Gordon Webster, *Partial Differential Equations of Mathematical Physics*, G. E. Stechert and Co., New York, 1927.

27. H. Bateman, *Partial Differential Equations of Mathematical Physics*, Cambridge University Press, The Macmillan Company, New York, 1932.

28. H. Levy and E. A. Baggott, *Numerical Studies in Differential Equations*, Watts, London, 1934.

29. R. V. Southwell, *Relaxation Methods in Engineering Science; A Treatise on Approximate Computation*, Oxford University Press, 1940.

30. Murray F. Gardner and John L. Barnes, *Transients in Linear Systems*, Vol. I, *Lumped-Constant Systems*, John Wiley and Sons, New York, 1942.

31. Harry F. Olson, *Dynamical Analogies*, D. Van Nostrand Company, Inc., New York, 1943.

32. Horace Lamb, *Hydrodynamics*, Cambridge University Press, 1932.

33. A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, Cambridge University Press, Fourth Edition, 1927.

34. Ivan S. Sokolnikoff, *Mathematical Theory of Elasticity*, McGraw-Hill Book Company, Inc., New York, 1946.

35. Samuel Glasstone, *Theoretical Chemistry*, D. Van Nostrand Company, Inc., New York, 1944.

36. Gerhard Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules*, D. Van Nostrand Company, Inc., New York, 1945.

37. Hendrik W. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand Company, Inc., New York, 1945.

38. LeRoy A. MacColl, *Fundamental Theory of Servomechanisms*, D. Van Nostrand Company, Inc., New York, 1945.

39. A. B. Basset, *An Elementary Treatise on Hydrodynamics and Sound*, Deighton Bell and Company, London, 1900.

40. A. S. Ramsey, *A Treatise on Hydromechanics*, Part II, *Hydrodynamics*, G. Bell and Sons, Ltd., London, 1913.

41. L. M. Milne-Thomson, *Theoretical Hydrodynamics*, Macmillan and Company, Ltd., London, 1938.

42. H. Glauert, *The Elements of Aerofoil and Airscrew Theory*, Cambridge University Press, The Macmillan Company, New York, 1943.

43. Leonard Bairstow, *Applied Aerodynamics*, Longmans, Green and Co., Ltd., London, New York, Toronto, Second Edition, 1939.

44. Julius Adams Stratton, *Electromagnetic Theory*, McGraw-Hill Book Company, Inc., New York, 1941.

45. R. S. Burington and C. C. Torrance, *Higher Mathematics: With Applications to Science and Engineering*, McGraw-Hill Book Company, Inc., New York, 1939.

46. P. W. Bridgman, *Dimensional Analysis*, Yale University Press, New Haven, 1922.

47. George Boole, *Collected Logical Works; The Law of Thought*, Vol. 2, 1854, Open Court Publishing Company, Chicago, 1916.

INDEX